

Study on Language Complexity of English Written Expression in Senior Three

Xue Xiao

China West Normal University, Nanchong City, Sichuan, 637000, China

Abstract

The purpose of this paper is to investigate whether foreign language writings of senior three at different proficiency levels can show distinguishing features in indices of language complexity. 42 senior three students in Qiuxi Senior High school are chosen as the research participants and their foreign language writings are considered as the research material. In this study, the author studies 14 lexical complexity indices and 14 syntactic complexity indices by using L2 Lexical Complexity Analyzer(LCA) and L2 Syntactic Complexity Analyzer(L2SCA), finding that there are no significant differences in most lexical complexity indices and all the syntactic complexity indices during the advanced senior three students and intermediate ones.

Keywords

Language Complexity; Foreign Language Writings; LCA; L2SCA.

1. Introduction

Lexical complexity and syntactic complexity are important indices to measure the quality of foreign language writings. In recent years, lexical complexity and syntactic complexity have attracted extensive attention at home and abroad, and have achieved rapid development. But there are still some problems in the existing research. First of all, there are some problems in the study of lexical complexity and syntactic complexity, such as the lack of syntactic indices. Secondly, most of the previous studies on language complexity focused on the University, and relatively few on the high school. In addition, most of the previous studies only focus on lexical complexity or syntactic complexity, and few of them combine them.

The grammar and vocabulary of senior three students are relatively stable, and their English writing levels are higher than those of senior one and senior two students. Therefore, this study chooses senior three students as the research participants. By analyzing the compositions' language complexity of the advanced senior three students and the intermediate ones, this study aims to understand the characteristics of different levels of students in the aspect of language complexity, so as to provide effective guidance for improving senior high school students' writing level, enrich the research on lexical and syntactic complexity.

2. Literature Review

2.1. Lexical Complexity

Lexical complexity refers to the scope and depth of learners' vocabulary knowledge (Bao Gui, 2011). Bao Gui found that lexical complexity is the best way to distinguish the differences of vocabulary use between different groups. Even if the measurement methods are different, the development path is still linear. Bao Gui(2008) & Kate Wolfe-Quintero(1998) pointed out that lexical complexity included lexical variation(LV), lexical density(LD), lexical sophistication(LS) and lexical originality(LO). In this study, lexical variation(LV), lexical density(LD) and lexical sophistication(LS) were measured. Hyltenstam, K. (1998) pointed out that quantitative measures of lexical usage focused on lexical variation, lexical density and lexical sophistication.

Liu Donghong(2003) thought that the factors that affected the quality of writing included text length, T-unit length and the amount of the first type words.

In this study, the author studies the differences during different writing quality and figures out which indices of lexical complexity are significantly different during good compositions and relatively bad ones by using L2 Lexical Complexity Analyzer(LCA). Lu (2010) summarized 25 vocabulary measurement indices and designed a free online vocabulary and syntax analysis system (<http://aihaiyang.com/synlex/lexical/>), based on previous studies. In this study, three aspects including lexical sophistication, lexical variation and lexical density, and 14 indices including lexical density(LD), lexical sophistication-I (LS1), lexical sophistication-II (LS2), Verb sophistication-I (VS1), verb sophistication-II (VS2), number of different words (NDW), type/token ratio (TTR), verb variation-1 (VV1), lexical word variation (LV), verb variation-II (VV2), noun variation (NV), adjective variation (AdjV), adverb variation (AdvV) and Modifier variation (ModV) were studied.

2.2. Syntactic Complexity

Syntactic complexity (also called syntactic maturity or linguistic complexity) refers to the range of forms that surface in language production and degree of sophistication of such forms (Ortega, 2003). Lu (2010) designed a free online vocabulary and syntax analysis system (<http://aihaiyang.com/synlex/lexical/>) in order to analyze the syntactic complexity of second language learners. Lu X. & Xu Qi (2016) pointed out that L2 Syntactic Complexity Analyzer(L2SCA) can automatically analyze English writing texts with 14 commonly used syntactic complexity measurement indices, which solved the bottleneck of data analysis in the research of syntactic complexity and effectively promoted the progress of the frontier research of syntactic complexity in second language writing. Lei Lei (2017) made a comparative study of syntactic complexity between English learners and native speakers by using L2SCA, finding that there were differences in the syntactic complexity of different English users, and some of the measurement indices could not reflect the genre characteristics of written language, or may be negatively related to the quality of writing. Pei Lixia (2020) studied the influence of different writing environments on the syntactic complexity of Chinese EFL learners' writing by using L2 Syntactic Complexity Analyzer(L2SCA) and Manual Marking.

In this study, the author used L2SCA to study syntactic complexity and chose 14 indices from Lu, including mean length of clause (MLC), mean length of sentence (MLS), mean length of T-units (MLT), clauses per sentence (C/S), clauses per T-unit (C/T), complex T-units per U-unit (CT/T), dependent clauses per clause (DC/C), dependent clauses per T-unit (DC/T), coordinate phrases per clause (CP/C), coordinate phrases per T-unit (CP/T), T-units per sentence (T/S), complex nominals per clause (CN/C), complex nominals per T-unit (CN/T) and verb phrases per T-unit (VP/T).

3. Method

3.1. Research Questions

In this study, two questions are explored:

Research Question1: Are there significant differences in lexical complexity between senior three students in the advanced group and in the intermediate group? What are the main indices of these differences?

Research Question2: Are there significant differences in syntactic complexity between senior three students in the advanced group and in the intermediate group? What are the main indices of these differences?

3.2. Participants and Material

42 senior three students, registered at Qiuxi Senior High school, participated in this study, including 21 senior three students who are more successful than others and 21 senior three students who are less. Chinese is their mother tongue and first language. Meanwhile, English is their foreign language. These 42 participants were chosen by the researcher randomly. The material, practical writing, in this study is from the monthly exam questions. Practical writing is the focus of composition teaching in senior high school and college entrance examination. In this test, participants were required to write articles about the theme of "Saving Food Is Our Duty". Besides, the number of words are around 100.

3.3. Research Instruments

In this study, the author uses L2 Lexical Complexity Analyzer(LCA) and L2 Syntactic Complexity Analyzer(L2SCA) to measure lexical and syntactic complexity. There are some reasons why these two instruments are used: Firstly, these two instruments are free and easy to use; Secondly, some researchers have used the instruments and proved that they are effective.

3.4. Research Producers

Firstly, the compositions of senior three four classes were recorded into word file by voice input, which avoided the influence of writing to the scores. An English teacher with rich teaching experience and a graduate student were asked to score after fully understanding the scoring rules. Then, the average score of the two was obtained as the final score. If the score difference is more than 3, then invite another English teacher to score, and the final score is the average of two close scores.

Secondly, divide students' writing level. The scores of the test were used as the standard to divide the writing level. The full score was 25, 20-25(including 20) was the advanced group, and 15-20 was the intermediate group. 21 participants in the advanced group and 21 participants in the intermediate group were selected.

Thirdly, the L2 syntactic complexity analyzer (Lu, 2010) and L2 lexical complexity analyzer (Lu, 2012) were used to analyze the 42 articles. The 42 students' compositions here were collected with the teachers' permission and without the students' knowledge.

Finally, SPSS was used to analyze the data obtained from LCA and L2SCA.

3.5. Data Analysis

To examine whether there exist significant differences in language complexity between senior three students in the advanced group and in the intermediate group, a series of statistic analysis was conducted by using SPSS 20.

To learn whether there are significant differences during the advanced and intermediate students, two-sample Mann-Whitney U was used. By analyzing the lexical and syntactic data, the homogeneity of variance did not exist. Therefore, two independent sample t-test couldn't be used. Then, two-sample Mann-Whitney U were chosen. 95% was chosen as Confidence Interval. If Asymp. Sig. is less than 0.05, there will exist significant differences during two sets of data.

From the data results, only the Asymp. Sig. of lexical density (LD), 0.012, is less than 0.05, which shows that there exists a significant difference in lexical density during the advanced senior three students and intermediate ones. However, there are no significant differences in other lexical indices during both these two kinds of students.

By using SPSS 20, the average, standard deviation and variance of lexical complexity indices are calculated and showed in Table 2.

4. Results

4.1. Lexical Complexity

According to Mann-Whitney U test, the results of lexical complexity indices data are showed in Table 1.

Table 1. Mann-Whitney U Results of Lexical Complexity Indices

Measure	Mann-Whitney U	Wilcoxon W	Z	Asymp.Sig.(2-tailed)
LD	122.5	353.5	-2.518	0.012
LS1	203	434	-0.442	0.659
LS2	161.5	392.5	-1.497	0.135
VS1	188	419	-0.853	0.394
VS2	185	416	-1.042	0.297
NDW	152	383	-1.725	0.085
TTR	197.5	428.5	-0.582	0.56
VV1	199	430	-0.542	0.588
LV	150	381	-1.802	0.072
VV2	204.5	435.5	-0.404	0.687
NV	152	383	-1.727	0.084
ADJV	210	441	-0.269	0.788
ADV	186.5	417.5	-0.862	0.388
MODV	182	413	-0.973	0.331

Table 2. Descriptive Statistics of Lexical Complexity Indices (a)

Measure	Average		SD		Variance	
	Advanced (N=21)	Intermediate (N=21)	Advanced (N=21)	Intermediate (N=21)	Advanced (N=21)	Intermediate (N=21)
LD	0.522	0.526	0.039	0.050	0.002	0.003
LS1	0.124	0.132	0.057	0.036	0.003	0.001
LS2	0.131	0.147	0.040	0.030	0.002	0.001
VS1	0.052	0.040	0.053	0.051	0.003	0.003
VS2	0.126	0.087	0.197	0.171	0.039	0.029
NDW	72.143	66.714	10.145	7.676	102.929	58.914
TTR	0.573	0.585	0.050	0.051	0.002	0.003
VV1	0.706	0.696	0.121	0.092	0.015	0.009
LV	0.700	0.687	0.080	0.055	0.006	0.003
VV2	0.226	0.221	0.043	0.053	0.002	0.003
NV	0.636	0.591	0.079	0.092	0.006	0.009
ADJV	0.114	0.110	0.028	0.041	0.001	0.002
ADV	0.102	0.095	0.034	0.032	0.001	0.001
MODV	0.217	0.204	0.044	0.037	0.002	0.001
LD	0.526	0.526	0.050	0.050	0.003	0.003

4.2. Syntactic Complexity

According to Mann-Whitney U test, the results of lexical complexity indices are showed in Table 3.

Table 3. Mann-Whitney U Results of Syntactic Complexity indices

Measure	Mann-Whitney U	Wilcoxon W	Z	Asymp.Sig.(2-tailed)
MLS	191	422	-0.742	0.458
MLT	205	436	-0.39	0.697
MLC	190	421	-0.767	0.443
C/S	215.5	446.5	-0.126	0.9
VP/T	203.5	434.5	-0.428	0.669
C/T	201.5	432.5	-0.478	0.632
DC/C	199.5	430.5	-0.529	0.597
DC/T	210	441	-0.264	0.791
T/S	201.5	432.5	-0.538	0.59
CT/T	216.5	447.5	-0.101	0.92
CP/T	192	423	-0.72	0.472
CP/C	198.5	429.5	-0.555	0.579
CN/T	202	433	-0.466	0.642
CN/C	182.5	413.5	-0.958	0.338

Table 4. Descriptive Statistics of Lexical Complexity indices (b)

Measure	Average		SD		Variance	
	Advanced (N=21)	Intermediate (N=21)	Advanced (N=21)	Intermediate (N=21)	Advanced (N=21)	Intermediate (N=21)
MLS	12.060	11.111	3.606	2.725	13.002	7.424
MLT	11.463	10.765	3.425	2.359	11.732	5.563
MLC	8.190	7.492	2.643	1.416	6.988	2.006
C/S	1.511	1.483	0.387	0.206	0.150	0.043
VP/T	1.981	1.976	0.602	0.350	0.362	0.122
C/T	1.442	1.441	0.385	0.192	0.148	0.037
DC/C	0.292	0.285	0.141	0.104	0.020	0.011
DC/T	0.468	0.426	0.336	0.207	0.113	0.043
T/S	1.053	1.030	0.095	0.059	0.009	0.003
CT/T	0.380	0.355	0.235	0.154	0.055	0.024
CP/T	0.150	0.159	0.144	0.108	0.021	0.012
CP/C	0.114	0.114	0.126	0.078	0.016	0.006
CN/T	1.237	1.079	0.576	0.333	0.332	0.111
CN/C	0.882	0.748	0.437	0.208	0.191	0.043

From the data results, all the Asymp. Sig. data are more than 0.05, which shows that there are no significant differences in syntactic complexity data during senior three students in the advanced group and in the intermediate group.

By using SPSS 20, the average, standard deviation and variance of syntactic complexity indices are calculated and showed in Table 4.

5. Conclusion

5.1. Major Findings

For lexical complexity, only there a significant difference in lexical density (LD) between senior three students in the advanced group and in the intermediate group. For other lexical complexity indices, there are no significant differences between those two groups. Maybe, phrasal collocation and accuracy of words couldn't be analyzed, so there are no significant differences in most lexical complexity indices. The average of most lexical complexity indices in advanced level are a little higher than that in middle level. Only, the average of number of different words(NDW) between these two groups is extremely different, and the average of NDW in the advanced group is higher than that in the intermediate group. Most of standard deviation and variance of lexical complexity indices between these two levels are close. However, the standard deviation and variance of number of different words(NDW) between these two groups differentiate greatly and those in the advanced group are far higher than those in the intermediate group.

For syntactic complexity indices, there are no significant differences between those two groups. The average, standard deviation and variance of most syntactic complexity indices in advanced group are a little higher than that in intermediate group. Only, the average, standard deviation and variance of mean length of clause (MLC), mean length of sentence (MLS) and mean length of T-units (MLT) between these two groups are quite different and those in advanced group are higher than those in intermediate group. Maybe, the advanced students pay more attention to the accuracy and fluency, so there are no significant differences in most language complexity indices during the advanced group and intermediate group.

5.2. Limitations and Suggestion

There are some limitations in this study. For one thing, the number of the research participants is a little small. For another, the indices of language complexity in this study can't represent language complexity completely. In the future research, bigger samples and more language complexity indices should be considered. In addition, the factors influencing the significant differences of language complexity should be explored.

References

- [1] Bao Gui, A Multi Dimensional Study on the Development of L2 Learners' Lexical Richness in Compositions [J].Technology Enhanced Foreign Language Education, 2008, (5):38-44.
- [2] Bao Gui, A Study on the Lexical Complexity of English Learners with Different Curriculum Levels [J]. Journal of PLA University of Foreign Languages, 2011, (4) : 55-60.
- [3] Hyltenstam, K. Lexical Characteristics of Near-Native Second-Language Learners of Swedish[J]. Journal of Multilingual and Multicultural Development, 1998, 9: 67-84.
- [4] Kim, Ji-Young. Predicting L2 Writing Proficiency Using Linguistic Complexity Measures: A Corpus-Based Study. English Teaching, 2014, 69(4), 27-51.
- [5] Liu Donghong. The Role of Vocabulary in English Writing[J]. Modern Foreign Language(Quarterly), 2003, 26(2):180-187.

- [6] Lei Lei. Syntactic Complexity of Chinese EFL Learners' Academic Writing[J]. Journal of PLA University of Foreign Languages, 2017, 40(5):1-9.
- [7] Lu, X. Automatic Analysis of Syntactic Complexity in Second Language Writing. International Journal of Corpus Linguistics, 2010, 15: 474-496.
- [8] Lu, X. The Relationship of Lexical Complexity to the Quality of ESL Learner's Oral Narratives[J]. The Modern Language Journal, 2012, 190-208.
- [9] Lu X., Xu Qi. L2 Syntactic Complexity Analyzer and Its Application in L2 Writing Research[J]. Foreign Language Teaching and Research(bimonthly), 2016, 48(3): 409-420.
- [10] Ortega, L. Syntactic Complexity Measures and Their Relationship to L2 Proficiency: A Research Synthesis of College-Level L2 writing [J].Applied Linguistics, 2003,(24) : 492-518.
- [11] Pei Lixia. The Influence of Different Writing Environments on the Syntactic Complexity of Chinese EFL Learners' Writing[J]. Journal of Heilongjiang Institute of Teacher Development, 2020, 39 (1): 128 -130.
- [12] Wolfe-Quintero, K., Inagaki, S. & Kim, H.-Y. Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity. Honolulu, HI: University of Hawaii Press, 1998.