A Study on the Syntactic Complexity of Writings of Non-English Major Learners

Xin Gong
(Chengdu Normal University, Chengdu, Sichuan 611130, China)

Abstract: This paper collects students' compositions from two levels of university English tiered teaching A,B, in an attempt to examine the syntactic complexity of students' writing in terms of several dimensional indicators, such as the length of linguistic output, the number of subordinate structures, the number of parallel structures and the complexity of phrases, etc. The results show that the higher the level of English proficiency, the higher the syntactic complexity of students; and comparing the writing of native speakers and Chinese learners', the results show that the syntactic complexity of Chinese learners differs significantly in terms of linguistic output, the use of specific phrases, and the overall structure of the sentences.

Keywords: syntactic complexity

Introduction

Firstly, we introduce the definition of syntactic complexity, which refers to syntactic maturity or linguistic complexity, the variation and complexity of linguistic forms in language output (Bulte & Housen 2014; crossley & McNamara 2014; Norris & Ortega 2009; Ortega 2003,2012). In second language learning, syntactic complexity is considered an important indicator for judging learners' language development (Lu 2011; Norris & Ortega 2009). In linguistics, syntactic complexity refers to the degree of complexity of the syntactic structure and form of a sentence. The syntactic complexity of a sentence can be determined by analyzing several aspects of its sentence constituents, grammatical structure, sentence length, and semantic complexity.

1. Current status of research

Syntactic complexity has been studied abroad in a variety of fields, including linguistics, psychology, education, and computer science. The following are some of the current status of foreign research on syntactic complexity:

Definition and measurement of syntactic complexity: many foreign scholars have proposed different definitions and measurements of syntactic complexity, such as Chomsky's theory of generative grammar, Biber's theory of linguistic variants and corpora, Hawkins' theory of thought load, etc. Research by Wei et al. (2021): they analyzed the syntactic complexity for different sentence types and proposed a syntactic complexity measure based on attention mechanism. Their study showed that the measure can accurately predict the syntactic complexity of different sentence types.

Factors influencing syntactic complexity: foreign scholars have conducted in-depth research and studies on the factors influencing syntactic complexity, including sentence length, sentence structure, language difficulty, contextual information, and language users' attention and working memory.Sorace, Preston-Dunlop, and Howell (2018) used ERP techniques and explored learners' sensitivity to complex language structure, and how sensitivity to linguistic structure is affected by syntactic complexity and language acceptability.

Applications in natural language processing: Syntactic complexity has a wide range of applications in natural language processing, for example, in machine translation, speech recognition, and text classification, where the measurement and processing of syntactic complexity are important issues. Daelemans et al. (2020),they used deep learning methods to measure syntactic complexity and applied it to a text classification task. Their study shows that syntactic complexity is one of the factors that have a significant impact on text classification.

Research on language disorders and aphasia: Syntactic complexity also plays an important role in understanding language disorders and aphasia, and foreign scholars' research results in this area are related to areas such as cognitive rehabilitation therapy. Hwang & Lin (2020),they investigated the relationship between syntactic complexity and language comprehension using large-scale cognitive linguistic experiments. They found that the syntactic complexity of different sentence types had different effects on comprehension speed and accuracy.

These recent studies suggest that syntactic complexity remains an active area of research and is becoming more prevalent in areas such as natural language processing and neuroscience. These recent studies provide new perspectives and approaches to our understanding of the development and mechanisms of syntactic complexity in English learners. They not only help us to better understand the mechanisms of language acquisition, but also help us to improve language education and teaching.

In recent years, more and more domestic researchers have begun to pay attention to the study of syntactic complexity of English learners' compositions, and the following are some of the current status of domestic research in China:

Syntactic complexity measurement indicators: Lei Lei (2017) studied the syntactic complexity of English learners and native speakers, and used a total of 10 syntactic complexity indicators. The results found that Chinese ELLs had shorter sentence output lengths and fewer subordinate structures in their writing relative to native speakers. In addition, by analyzing the corpus of two different genres of compositions in an English writing competition held between colleges and universities in one province,

The relationship between syntactic complexity and learners' language proficiency: the relationship between syntactic complexity and second language proficiency was explored by Xue Wu and Lei Lei (2018) using a meta-analysis. The results showed that second language

level may have an impact on learners' syntactic complexity if different measures are selected for testing.

Exploring the effects of individual variables on syntactic complexity. The study by Xue Wu (2017) examined the effects of journal articles from different disciplines on syntactic complexity, while Xiaofang Chen (2017) focused on the effects caused by automated writing scoring systems on syntactic complexity. The findings demonstrate that syntactic complexity varies widely across disciplines, which may have a significant impact on the readability of texts.

2. Research questions

1.To investigate how the second language level of non-English major learners correlates with syntactic complexity?

2.To detect the differences in the syntactic complexity measures of college English learners A and B by comparing and analyzing their writing at two different proficiency levels, and to analyze the reasons for these differences.

3.To compare the writing of Chinese non-English learners with that of native speakers to find out the differences in their syntactic complexity measures and to explain the reasons for the differences from the perspective of language acquisition.

3. Research methodology

1. Corpus introduction:

In this study, we collected the argumentative essays from two semesters of college English (I) and college English (II) from four classes of non-English major learners(A and B)of different levels. Among them, Class A3 is majoring in Physics and Class B3 is also majoring in Physics; Class A10 is majoring in Chinese Language and Literature and Class B11 is also majoring in Chinese Language and Literature. The native speaker corpus was selected from undergraduate theses. The details of the corpus are as follows:

2. Syntactic Complexity Measurement Metrics and Tools

This study used the L2 Syntactic Complexity Analyzer (L2SCA) developed by Lu (2010) to measure the syntactic complexity of students' writing in several dimensions such as length of linguistic output, the number of subordinate structures, the number of parallel structures, and phrase complexity in order to assess and analyze differences in the quality of students' writing.

In this study, firstly, L2SCA was used to mark each student's writing to obtain the data of syntactic complexity index, and then SPSS 25 was used for comparative analysis to find out the differences in syntactic complexity index between students of A and B levels and then compare them with native speakers. Finally, these findings are interpreted using quantitative or qualitative methods to help readers gain a deeper understanding of the meaning behind these data.

4. Findings of the study

1. Firstly, the L2SCA was used to mark two levels of syntactic complexity indicators for non-English major learners, A and B, for descriptive statistics, as well as for the writing of native speakers. As shown in the figure:

Table 1 Corpus Statistics

Corpus statistics		
Corpus	Number of texts	Number of words
A3 first time collected	55	8650
A3 second time collected	59	8235
B3 first time collected	61	101811
B3 second time collected	58	9398
A10 first time collected	53	7531
A10 second time collected	35	5891
B11 first time collected	63	10101
B11 second time collected	51	7030
Native speakers	20	170100

Table 2 The Results of Syntactic Complexity Analyzer

syntactic complexity indicators		A3 first time collected		A3 second time collected		A10 first time collected		A10 second time collected		B3 first time collected		B3 second time collected		B11 first time collected		B11 second time collected		Native speakers	
		mean	Sd	mean	Sd	mean	Sd	mean	Sd	mean	Sd	mean	Sd	mean	Sd	mean	Sd	mean	Sd
The output of linguistic length	MLS	18.89	9.56	22.8	13.53	25.64	12.13	28.28	35.84	25.09	26.25	20.12	6.27	15.87	6.59	22.65	15.99	28.02	7.88
	MLT	15.27	6.08	18.9	10.86	22.61	10.1	19.04	11.25	17.33	8.72	17.73	4.79	14.46	6.14	19.04	10.39	26.1	7.7
	MLC	8.11	1.42	9.13	2.03	9.63	1.8	10	2.18	9.76	2.22	11.23	2.13	8.01	1.3	10.5	2.38	15.45	3.02
The number of subordinate structure usage	CT_T	0.51	0.21	0.53	0.22	0.56	0.24	0.45	0.21	0.44	0.21	0.43	0.18	0.47	0.2	0.4	0.21	0.46	0.18
	DC_T	0.83	0.6	0.92	0.97	1.2	0.86	0.81	0.9	0.72	0.7	0.53	0.25	0.7	0.6	0.6	0.64	0.72	0.54
	C_T	1.91	0.77	2.1	1.28	2.43	1.27	1.94	1.28	1.85	1.08	1.59	0.34	2	1.38	1.78	0.85	1.76	0.73
	DC_C	0.4	0.14	0.38	0.12	0.45	0.14	0.36	0.13	0.35	0.14	0.32	0.1	0.35	0.11	0.3	0.12	0.37	0.12
The number of Parallel structure usage	CP_T	0.36	0.25	0.33	0.29	0.54	0.33	0.65	0.56	0.44	0.36	0.62	0.33	0.32	0.25	0.55	0.37	0.75	0.22
	T_S	1.25	0.43	1.24	0.58	1.14	0.27	1.2	0.6	1.5	1.73	1.14	0.19	1.1	0.15	1.22	0.67	1.08	0.08
	CP_C	0.19	0.11	0.17	0.12	0.24	0.13	0.35	0.21	0.25	0.17	0.39	0.2	0.19	0.14	0.32	0.16	0.47	0.21
Specific Phrase Structure	VP_T	2.41	1	3	1.64	3.8	1.84	2.6	1.53	2.47	1.35	2.26	0.59	2.28	1.14	2.62	1.4	2.67	1.1
	CN_C	0.88	0.27	1.13	0.37	1.23	0.31	1.07	0.33	1.2	0.45	1.51	0.52	0.86	0.25	1.43	0.43	2.11	0.4
	CN_T	1.73	1.1	2.36	1.46	1.57	0.85	2.1	1.52	2.23	1.8	2.34	0.8	1.57	0.85	2.58	1.64	3.56	1.09
Overall sentence structure	C_S	2.42	1.5	2.6	1.76	2.76	1.47	2.46	2.38	2.68	2.89	1.8	0.46	2.03	1	2.17	1.73	1.89	0.75

As can be seen from the above table, there are significant differences between English proficiency levels A and B, and between Chinese English learners and native speakers in three areas: output of the linguistic length, specific phrase structure, and use of parallel structure, indicating that both English proficiency and differences in the use of native language have an impact on the measures of syntactic complexity. Therefore, the training and improvement of these syntactic complexity indicators should be focused on in English learning and teaching in order to improve learners' overall English proficiency.

It is worth noting that among these syntactic complexity indicators, Chinese learner A10 is higher than the average of native speakers in the use of VP_T for subordinate structures, indicating that the class is more competent in this area of application, which should be fully recognized and guided in the teaching.

2. Secondly, to investigate the differences in the syntactic complexity measures among different English learners, we used spss25 to conduct the Mann-Whitney test and find out the differences between English learners of different proficiency levels and native speakers. The specific results are detailed in the following tables:

Table 3 Mann-Whitney Test

syntactic complexity indicators		Groups Comparison			
		A3 vs B3	A10 vs B11	A10 vs Native speakers	B11 vs Native Speakers
The output of linguistic length	MLS	0.178	0.000	0.000	0.000
	MLT	0.037	0.000	0.000	0.000
	MLC	0	0.000	0.000	0.000
The number of subordinate structure usage	DC_T	0.12	0.000	0.835	0.069
	CT_T	0.09	0.047	0.497	0.119
	C_T	0.121	0.002	0.887	0.62
	DC_C	0.069	0.000	0.393	0.006
The number of parallel structure usage	CP_T	0.207	0.000	0.009	0.000
	T_S	0.363	0.85	0.815	0.767
	CP_C	0.043	0.01	0.007	0.001
Specific phrase structure	VP_T	0.537	0.000	0.093	0.152
	CN_T	0.011	0.000	0.000	0.000
	CN_C	0	0.000	0.000	0.000
Overall sentence structure	C_S	0.163	0.003	0.857	0.992

According to the Mann-Whitney test, if P<0.05, it means that there is a difference between the comparison groups. In terms of language output length indicators, there were differences between comparison groups A3 and B3, except the MLS which was 0.178>0.05. And for A10 and B11, the differences between them were highly significant (P=0.000); also, the differences between A10 and the native speakers were highly significant, while there were significant differences between B11 and the native speakers. In terms of subordinate clause usage, the difference between A10 and B11 was highly significant; while there was no significant difference between the comparison groups except for the difference between B11 and the native speakers on the DC_C indicator. It should be noted that the value of 0.393 on the DC_C indicator between A10 and the native speakers is higher than the contrastive value between the other groups. Therefore, we can assume that Chinese English learners do not differ significantly from native speakers in the use of subordinate clauses. In terms of the use of parallel structure, there were significant differences between the comparison groups except for A3 and B3 which showed no difference in the CP_T indicator, and the groups differed significantly in the CP_C indicator. There were no differences between the groups on the T_S indicator. For the usage of specific phrase structure, the data between the comparison groups showed that all values were less than 0.05 except for the value of VP_T indicator which was higher than 0.05, therefore, we can infer that there were significant differences between the groups in the usage of specific phrase structure. In terms of overall sentence structure, statistics from spss25 reveal that there are no significant differences between the comparison groups.

3. Based on the results of the study, it can be concluded that there are significant differences in the use of syntactic complexity among non-English major learners with different levels of English proficiency. In particular, the higher the level of English proficiency, the higher the syntactic complexity used by students in terms of length of linguistic output, the amount of parallel structure, and specific phrase structure, and vice versa. In terms of professional background, students in A10 and B11 were Chinese language and literature majors, while students in A3 and B3 were physics majors. Comparing the two sets of data, it can be found that there are differences between A3 and B3 in some of the syntactic complexity indicators, such as the length of linguistic output and the use of specific phrase structures. While A10 and B11 differed in all indicators of syntactic complexity, can it be assumed that the syntactic complexity of Chinese language and literature majors is higher?

It was also found that there were differences in the use of syntactic complexity when comparing between Chinese non-English major learners and native speakers, highlighted in the use of linguistic output length, specific phrase structure, and parallel structure. This suggests that our English language teaching focuses more on the use of subordinate clauses, that is, on the teaching of grammar. However, we still need to continue to improve and strengthen the teaching of language length, the use of parallel structures, and the use of specific phrase structures.

In summary, this study explores the syntactic complexity of non-English major learners and also reveals the shortcomings of English language teaching. These findings provide some reference and guidance for our future English educational teaching and research, as well as specific suggestions.

Reference:

[1]Chen, S. F.. The effect of automatic writing scoring on syntactic complexity[J]. Journal of Heilongjiang Institute of Technology: Comprehensive Edition. 2017, 011: 122-125.

[2]Lei Lei. A study on syntactic complexity of academic writing of Chinese English learners[J]. Journal of PLA Foreign Language Institute, 2017, 40(05): 1-10+159.

[3]Lu S. F., Xu Q. Two-statement method complexity analyzer and its application in second language writing research [J]. Foreign Language Teaching and Research, 2016, 48(03):409-420.

[4]Wu Xue, Lei Lei. A meta-analysis of second language level and syntactic complexity research[J]. Modern Foreign Languages, 2018, 41(04): 481-492.

[5]Wu Xue. Research on syntactic complexity and text readability of Chinese scholars' international journal papers[J].Journal of the PLA Foreign Language Institute, 2017, 40(05):11-19+159.

Author: Gong Xin, female, lecturer, born in May 1982, master's degree student, currently working in Chengdu Normal University, with research interests in foreign linguistics and applied linguistics.

Note: This paper is the conclusion of a general project (Project No. CS20SB08) of Chengdu Normal University.