ISSN 2522-6398 Vol. 6, Issue 16: 124-128, DOI: 10.25236/FER.2023.061621

A Comparative Study of the Syntactic Complexity of English Writing for High School Students

Xiaomin Deng[1]

[1]*College of Education, China West Normal University, Nanchong, 637000, China*

Abstract: *Based on the self-built corpus and LOCNESS corpus, combined with L2 syntactic Complexity measurement index, SPSS, and other software are utilized to explore the commonality and difference in syntactic complexity between English language learners and native speakers with a large-scale corpus of compositions. Concluding the study, English language learners were significantly lower than native speakers in seven of these measures. Although there is not any statistically significant difference in the average length measure of T-unit, there is a large numerical difference. The results have a certain significance for teaching and learning in second language writing.*

Keywords: *syntactic complexity, English language learners, native speakers, second language writing*

1. Introduction

The National English Curriculum Standards for General Senior High Schools clearly state that the core literacy to be developed in the English curriculum includes linguistic competence, cultural awareness, thinking quality, and learning ability, with linguistic competence being a fundamental element of the core literacy.[1]Linguistic competence is a multidimensional compound concept, and linguistic complexity, accuracy, and fluency are considered to be indispensable elements for measuring linguistic competence and writing proficiency.[2]In this context, the syntactic complexity of compositions is compared with that of native speakers to explore the commonalities and differences between the two groups of learners, to synthesize the syntactic features of English learners' written language, and to assess their language development and ability to use language in practice, to provide new insights into English writing teaching practice.

1.1 Syntactic Complexity in Second Language Writing

Syntactic complexity, also known as syntactic maturity or linguistic complexity, refers to the range of variation in syntactic forms and the degree of formal complexity in linguistic output.[3-6]In second language writing teaching practice and research, syntactic complexity is considered to be one of the keys to assessing second language learners' language proficiency and writing quality. Qiu argues that syntactic complexity measures can reflect second language learners' syntactic knowledge reserves and their ability to use language, which is related to the integration of conceptual and knowledge resources in the process of language output.[7] Whereas writing ability is a comprehensive reflection of second language proficiency. It is particularly important to comprehensively investigate the syntactic complexity of second language learners' writing.[8]

In addition to the relationship between syntactic complexity and second language writing, few papers have so far specifically explored the common and differential features of syntactic complexity between high school English learners and native speakers of the same age. Ai&Lu utilized 10 measures to compare the syntactic complexity of Chinese college students' writing with that of native speakers. The study found that most of the syntactic complexity measures, except for the subordinate structure dimension, increased with language proficiency, but the subordinate structure used by second language learners was significantly lower than that used by native speakers.[9] Based on the corpora, Huiping Zhang and Siyu Zhang investigated the development trend of the syntactic useability of secondary school students based on the language exposure hypothesis.[10]It was found that the use of complex nominal gradually increased, indicating that their syntactic forms were gradually becoming more academic.[10]The former study focused on the differences in syntactic complexity among domestic and foreign university students, while the latter discussed the differences in syntactic complexity among junior high school students in China, Hong Kong, and Taiwan. However, few studies explored

ISSN 2522-6398 Vol. 6, Issue 16: 124-128, DOI: 10.25236/FER.2023.061621

the comparison in syntactic complexity between high school students and native speakers.

We believe that the commonalities and differences between the syntactic complexity characteristics of English language learners' second language writing and those of their native speakers can help teachers understand where the current differences of students lie so that they can think about how to improve students' writing skills in terms of different dimensions of syntactic structure, which is the purpose of this study, to have some reference value for second language writing teaching practice.

1.2 Measuring L2 syntactic complexity

Syntactic complexity covers a large variety of features, including length of the production unit, Amount of subordination, Amount of coordination, and Degree of phrasal per T-unit which gauges syntactic complexity in second language learners' writing development.[6][11]In terms of diversity, syntactic complexity covers four dimensions as well as multiple measures of syntactic structure, and in terms of complexity, it includes subordinate clause and phrase complexity as a way to present the complexity of syntactic structure.[12] Lu Xiaofei and Xu Qi argue that a systematic investigation of the relationship between syntactic complexity and other dimensions in language development and second language writing requires a large collection of measures to analyze a large amount of second language writing research to data to help us gain insight into the status and role of syntactic complexity in the development of second language and second language writing.[13] The syntactic complexity analysis tool used in this study is Lu Xiaofei's L2SCA, which extracts 14 syntactic complexity measures for the English composition of English learners and native speakers.[14]

2. Method

2.1 Description of the Corpora

The corpus of native speakers sampled in this study is the *Louvain Corpus of Native English Essays* (LOCNESS), which includes 204 essays written by high school students in the UK and 232 essays written by college students in the US, totaling 324,304 words. In addition, we selected 10 essays on the topic of technology from its sub-corpus British A Level. The English learner corpus was collected from the writing texts of students in a high school in Nanchong, including 23 topic essays, from which a total of 40 essays were selected based on the need of the study, with the same topic and the same genre as the native speakers, and all of them were time-limited, and both groups were randomly sampled to form a small corpus. To ensure the comparability of syntactic complexity between the two groups, it is necessary to ensure that the word tokens and word types of the sampled corpus is the same. As shown in Table 1, the number of compositions in the two corpora differed greatly due to the large difference in the average length of compositions, whereas the syntactic complexity measures selected in this study refer to the average length of output units or the ratio of the frequencies of two syntactic structures. Given that the length of the compositions did not affect the results of the syntactic complexity measures.[15]

Table 1: Basic information of the sampled corpora

Corpora	Word Tokens	Word Types	Numbers of essays
LOCNESS-1	5077	1190	10
NCYZ-1	5346	701	40

2.2 Research questions

The present research aims to systematically examine differences in syntactic complexity in English writing between English language learners and native speakers. We seek to answer the following two questions:

Research question 1: What is the overall syntactic complexity used by English language learners?

Research question 2: What are the common and differential features of syntactic complexity in English writing between native speakers and English learners?

ISSN 2522-6398 Vol. 6, Issue 16: 124-128, DOI: 10.25236/FER.2023.061621

2.3 Data analysis

The data analysis can be divided into the following three steps: firstly, based on the Haiyang Ai website, the values of 14 syntactic complexity measures of the written texts in the 50-composition corpus, including length of the production unit, Amount of subordination, Amount of coordination, and Degree of phrasal per T-unit, which were obtained by L2SCA. Secondly, English learners and native speakers were compared as a whole, and the article used independent sample t-tests to compare the differences in the use of the 14 syntactic complexity measures on their composition data. Finally, a discussion and relevant conclusions are drawn based on the differences in the different measures.

3. Results and Discussion

3.1 English learners' overall syntactic level of syntactic complexity

As set out in table 2, the current study found that English learners were overall lower than native speakers on the syntactic complexity measures. Specifically, English learner composition has 12 measures of syntactic complexity lower than native speakers, most notably length of production unit(mean length of clause MLC, mean length of sentence MLS, mean length of T-unit MLT), followed by the amount of subordination (dependent clauses per clause DC/C, complex T-units per T-unit CT/T, clauses per T-unit C/T, clauses per sentence C/S), degree of phrasal sophistication(complex nominal per clause CN/C, complex nominal per T-unit CN/T) and finally amount of coordination(coordinate phrases per clauses CP/C, coordinate phrases per T-units CP/T).

The number of dependent clauses per T-unit (DC/T) and T-units per sentence (T/S) were used at a higher rate than native speakers. The present study demonstrates that English learners use more simple syntactic structures and use too little of more complex structures (subordinate clauses, phrases, and coordination structures), indicating a larger gap between English learners and native speakers. This finding is consistent with Siyu Zhang's and Huiping Zhang's research on the syntactic complexity of junior high school students' compositions.[15]

Table 2: Comparison of syntactic complexity indicators of written texts between English learners and native speakers

measures	Code	Native speaker Mean(SD)	Learners Mean(SD)	t	Sig
Length of production unit					
Mean length of clause	MLC	17.261(2.912)	9.335(2.137)	9.738	.000
Mean length of sentence	MLS	21.180(3.691)	12.380(5.731)	4.604	.000
Mean length of T-unit	MLT	21.177(3.690)	13.953(8.515)	2.606	.012
Amount of subordination					
Dependent clauses per clause	DC/C	1.686(0.217)	0.303(0.115)	27.998	.000
Dependent clauses per T-unit	DC/T	0.366(0.093)	0.495(0.413)	-.972	.336
Complex T-units per T-unit	CT/T	1.236(0.179)	0.376(0.208)	11.991	.000
Clauses per T-unit	C/T	2,260(0,311)	1.498(0.745)	3.149	.003
Clauses per sentence	C/S	10.243(1.212)	1.339(0.523)	35.684	.000
Amount of coordination					
Coordinate phrases per clause	CP/C	0.331(0.125)	0.227(0.137)	2.188	.034
Coordinate phrases per T-unit	CP/T	0.507(0.154)	0.336(0.222)	.166	.026
T-units per sentence	T/S	0.633(0.223)	0.911(0.112)	-5.645	.000
Degree of phrasal sophistication					
Complex nominal per clause	CN/C	2.023(0.488)	0.856(0.314)	9.351	.000
Complex nominal per clause	CN/T	0.204(0.102)	1.304(1.031)	-3.344	.002
Verb phrases per T-unit	VP/T	2.063(0.233)	2.723(1.961)	-1.054	.297

3.2 Comparison of syntactic complexity indexes between the two groups of learners

3.2.1. Length of the production unit

From the results of the comparison of the syntactic measures of the length of the production unit, we know that the mean length of clause (MLC=17.261), mean length of sentences

ISSN 2522-6398 Vol. 6, Issue 16: 124-128, DOI: 10.25236/FER.2023.061621

(MLS=21.180), and mean length of T-unit(MLT=21.177)are significantly higher in the compositions of native speakers than English learners, and there are statistically significant differences in two of the length of production unit measures(MLC,p = .000, MLS,p = .000), and the measure that was not statistically significant difference was the mean length of T-units(MLT, p = .012), but still had a large numerical difference compared to native speakers. This tells us that at some level it can be indicated that native speakers generally have longer sentences in their compositions and that the three measures are different constructs, such as mean length of sentence, mean clause length, and mean subject-subordinate sentence length, measuring language length from different perspectives. This finding is consistent with Yan Sheng's conclusion that MLS and MLT measures are significantly higher than those of learners and that there is no significant difference in MLC.[16]

3.2.2. Amount of subordination

The comparison of the dimension of the amount of subordination revealed that the mean values of the four measures in this dimension were DC/C=.303, CT/T=.495, C/T=1.498, and C/S=1.339 for English language learners' written texts, which were all lower than the values of the corresponding measures for native speakers' compositions, with the measures of dependent clauses per clause(DC/C), complex T-unit per clause(CT/T) and clauses per sentence(C/S) being significantly higher than the English learners. The dependent clauses per T-unit (p = .336)and clauses per T-unit(p = .003)were not significantly different. Yan Sheng thought these measures as stable measures to distinguish the number of subordination structures in written texts between native speakers and English learners.[16]

3.2.3. Amount of coordination

The results of the comparison of the amount of coordination can be observed in the table. The mean values of the three syntactic measures for ELL compositions are CP/C= .227, CP/T=.336, and T/S=.911, and the value of T-units per sentence in the clause is significantly higher than that of the native speakers, which is contrary to the conclusion reached by Yan Sheng, thus verifying that this measure is the most reliable coordination to distinguish the composition of native speakers and English learners. Coordinate phrases per clause and coordinate phrases per T-unit were not significantly different from those of native speakers.

3.2.4. Degree of phrases sophistication

As shown in the table, the results of the degree of phrase sophistication comparison can be learned, and the mean values of the three measures of English learners' compositions are CN/C=.856, CN/T= 1.304, and VP/T=2.723, in which complex nominal per clause is significantly lower than that of native speakers (CN/C P=.000), and He Xinyi et al found on the exploration of syntactic features of the writing texts of high-proficiency of students. The writing texts of high-proficiency students were found to have a prominent performance in complex nominal per clause (CN/C), which was one of the most expressive syntactic features.[17] Regardless of the fact that this study did not distinguish between high and low proficiency. The comparison with native speakers shows that English learners as a whole are less likely to use complex nominal structures. The other two measures (CN/T, VP/T)were neither statistically significantly different nor differed much in value.

In summary, compared to native speakers, English language learners use more complex syntactic structures such as clauses and phrases as well little in favor of simple syntactic structures, and the lengths of the production unit is generally lower than those of native speakers. Foreign language teachers generally feel that English writing is a weak area for their students, but feel overwhelmed by how to help students improve their writing skills, based on the findings of the study, to provide references.[18]

4. Conclusion

Based on the self-constructed corpus and the LOCNESS British A Levels composition corpus, this study explored the commonalities and differences in syntactic use between English learners and native speakers in 12 of the 14 syntactic complexity measures, with 7 of them being significantly lower than native speakers. Based on the findings, this study has definite implications for writing instruction. First, it can help front-line teachers understand the gap in syntactic structure between English language learners and native speakers, and teachers can develop more effective and targeted writing instruction programs based on the current situation of English language learners' use of different dimensions of syntactic structure indicators. In terms of the number of subordinate

ISSN 2522-6398 Vol. 6, Issue 16: 124-128, DOI: 10.25236/FER.2023.061621

structures, teachers should emphasize the use of subordinate clauses and the transition from learning to writing, such as definite clauses, emphatic clauses, and inverted clauses. In terms of parallel structure, English learners are more stuck in the analogous application of and, but, and individual phrases, lacking in sentence variation. Teachers should convert their mindset to use parallel structure flexibly and pay attention to developing their ability to synthesize various sentence types to enhance sentence complexity. In terms of phrase complexity, which is more demanding for learners, teachers can provide targeted instruction according to students' writing levels or different score bands.

The number of samples made in this study was limited and the group was relatively fixed, so the adequacy of the sample's representatives depends on further testing and validation in subsequent studies. In the future, the differences in syntactic structure can be investigated by dividing the high and low score bands according to the authority ratings.

References

[1] Developed by the Ministry of Education of the People's Republic of China. (2003). English curriculum standards for general high school: experimental. People's Education Press.
[2] Liu Ligang & Miao Haitao. (2018). Theory and Measurement of Language Complexity. Foreign Language Studies (01), 52-55. doi:10.13978/j.cnki.wyyj. 2018.01.010.
[3] Bram Bulté and Alex Housen. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. Journal of Second Language Writing, 26pp. 42-65.
[4] Scott A. Crossley and Danielle S. McNamara. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. Journal of Second Language Writing, 26pp. 66-79.
[5] Norris J. M., & Ortega L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. Applied linguistics, 30(4), 555-578.
[6] Ortega L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. Applied linguistics, 24(4), 492-518.
[7] Qiu J. (2014). The Effect of Internet-based Read-Write Teaching on the Syntactic Complexity of English Writing. Modern Foreign Languages, 37(6), 846-54.
[8] Wu X. & Lei L. (2018). A meta-analysis of second language level and syntactic complexity studies. Modern Foreign Languages (04), 481-492.
[9] Ai H., & Lu X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. Automatic treatment and analysis of learner corpus data, 59.
[10] Zhang Huiping & Zhang Siyu. (2020). A study on the developmental characteristics of secondary school students' written syntax complexity. Foreign Language Studies (06), 60-66+73. doi:10. 13978/j.cnki.wyyj. 2020.06.009.
[11] Bulté B., & Housen A. (2012). Defining and operationalising L2 complexity. Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA, 32, 21.
[12] Mancilla R. L., Polat N., & Akcay A. O. (2017). An investigation of native and nonnative English speakers' levels of written syntactic complexity in asynchronous online discussions. Applied Linguistics, 38(1), 112-134.
[13] Lu S. F., & Xu Q. (2016). A two-speech method complexity analyzer and its application in second language writing research. Foreign Language Teaching and Research, 48(3), 409-420.
[14] Lu X. (2010). Automatic analysis of syntactic complexity in second language writing. International journal of corpus linguistics, 15(4), 474-496.
[15] S. Y. Zhang, & H. P. Zhang. (2019). A comparative study on the complexity of Chinese junior high school students' written grammar. Basic Foreign Language Education.
[16] Yan Sheng. (2022). A comparative study on the syntactic complexity of same-topic writing between Asian English learners and native English speakers. Shandong Foreign Language Teaching (05), 44-55. doi:10.16482/j.sdwy37-1026. 2022-05-005.
[17] He X. Y., Lai Y. W., Li L. Y., Zou X. L. & Zhang S. H. (2018). Exploring the syntactic features of English writing texts in high subgroups—Based on data mining of English writing texts on sentencecoolcritic.com. Modern Educational Technology (12), 74-79.
[18] Fang X. (2012). Teaching and Researching English Writing. Foreign language teaching and research press.