

Syntactic Complexity of EFL Chinese Students' Writing

Sue Wang¹ & Tammy Slater²

¹ School of Foreign Studies, Central University of Finance and Economics, Beijing, China

² Applied Linguistics and Technology Program, Iowa State University, Ames, IA, United States

Correspondence: Sue Wang, School of Foreign Studies, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing, China. E-mail: suewangcufe@126.com

Received: January 19, 2016 Accepted: February 3, 2016 Online Published: February 26, 2016

doi:10.5539/ells.v6n1p81

URL: <http://dx.doi.org/10.5539/ells.v6n1p81>

Abstract

Syntactic complexity as an indicator in the study of English learners' language proficiency has been frequently employed in language development assessment. Using the Syntactic Complexity Analyzer, developed by Lu (2010), this article collected data representing the syntactic complexity indexes from the writing of Chinese non-English major students and from the writing of proficient users of English on a similar task. The results indicate that there is a significant difference in the use of complex nominals, the mean length of sentences, and the mean length of clauses between the writings of EFL Chinese students and more proficient users. This study provides suggestions for EFL writing teaching, particularly writing at the sentence level.

Keywords: syntactic complexity, EFL Chinese, writing

1. Introduction

Language is considered an important skill to have in the current context of globalization. For Chinese students who learned English as a foreign language (EFL), the quality of their writing is an important index of their language proficiency development. Their writing development needs to be assessed from a wide range of indexes. Syntactic complexity as one of those indexes refers to "the range of forms that surface in language production and the degree of sophistication of such forms" (Ortega, 2003, p. 492). It is one of several important measures of the proficiency or development of language learners and plays an important role in language testing and evaluation.

Literature on ESL students' writing has highlighted the syntactic complexity issues related to L2 writing. Silva (1993) found that there are significant differences in terms of fluency, accuracy, and syntactic structure between the written texts of native speakers and second language speakers. Hinkel (2003), after an analysis of the academic texts by native and non-native English speakers in American universities, also found that L2 writers tend to overuse simple sentence structures. In order to better understand the syntactic complexity of language learners, a number of researchers have explored this issue; the following is a brief review of the studies in this area.

2. Literature Review

Researchers have in many decades investigated the syntactic complexity of language learners (e.g., Larsen-Freeman, 1978; Henry, 1996; Lu, 2010, 2011). These studies were mostly conducted through quantifiable complexity indexes that include length of production unit, sentence complexity, and the frequency of a range of sentence structures. Among these, the T-unit (Hunt, 1965), the shortest grammatical chunk of a sentence as a unit of analysis, is an important concept and index. Wolfe-Quintero et al. (1998) reviewed 39 articles on L2 writing discussing multiple indexes for accuracy, fluency and syntactic complexity. The authors found that mean length of T-unit, mean length of clause, mean number of clauses per T-unit and dependent clauses per clause are the best indexes to measure syntactic complexity. Besides the four indexes mentioned, mean length of sentence and mean number of T-units per sentence have also been included as indicators for syntactic complexity (Ortega, 2003).

In studies such as Ortega (2003) and Wolfe-Quintero et al. (1998), the measurement based on T-unit and clauses is often listed and widely accepted as an important index for language development. However, some studies have pointed out that more proficient language learners are not necessarily using more T-units or clauses. For instance, Rimmer (2006) argued that syntactic complexity should include phrasal features such as noun post-modifiers.

Taguchi et al. (2013) found in their studies that noun phrase modifiers (including preceding attributive adjective and prepositional phrase as post modifiers of nouns) can be an indicator of writing quality. Biber, Gray, & Poonpon (2011) further questioned the measurement of syntactic complexity through T-unit based indexes. The above-mentioned six indexes for measuring syntactic complexity are thus far away from being conclusively determined.

In China, some grammatical complexity studies about EFL Chinese students' syntactic complexity have been conducted but they have mostly focused on the study of vocabulary complexity (Bao, 2009). There are only a few studies that have addressed the topic of syntactic complexity. For instance, Qin & Wen (2007) explored the syntactic complexity of English majors in China and found that the students' length of T-unit and clause increased linearly as they advanced in their studies. Bao (2009) and Shen & Bao (2010) investigated the length and density of sentences. These authors found similar results in terms of length development in the students' writing in these studies. However, Bao also pointed out that in comparison to native English writers, English learners showed an inadequacy in their density index development. Xu et al. (2013) compared the length of T units and clauses, sentence density as reflected in embedded clauses which includes the ratios of clauses to T-units and of dependent clauses to clauses, as well as the syntactic structures covering independent and independent clauses, passives and reduced structures. They found that Chinese students differ significantly from native speakers both in terms of sentence length and density. The findings have suggested that Chinese students still need to improve and develop their abilities to use complex sentences.

Up to now, findings on syntactic complexity indexes as indicators of language development or proficiency have been inconsistent. While some researchers consider T-unit-based measures adequate for syntactic complexity, others have argued that there are other indexes that should be included. Therefore, it is of necessity to explore further the syntactic differences between ESL / EFL students' writing and that of proficient users. With this in mind, and to contribute to the literature in this area, the current study was designed to explore the syntactic complexity differences between Chinese learners of English and proficient users of English. On a more practical level, by describing more accurately the syntactic development of Chinese learners of English and their challenges and difficulties with syntactic complexity, instruction could be better designed to target those relevant areas. At the same time, the development of syntax is universal, and therefore the study can provide thoughts and insights for other ESL or EFL learners at the tertiary level, particularly on the sentence level.

The research questions of this study asks whether there are any syntactic complexity differences between EFL Chinese learners' writing in comparison with the more proficient English users and what are the differences if any?

3. Methodology

3.1 Data Sources

The data used in this study is from documents commonly known as personal statements (PS). As a required document for graduate admission in most universities, it is a way to demonstrate the applicants' writing level. The data collected for this study include personal statements written by EFL learners and personal statements written by English proficient users. According to the findings from Lu (2011), the types of tasks and writing time impact the syntactic complexity. Therefore, we chose a task that was similar for all writers and for which the writing time was not limited; thus it would be comparable to analyze the syntactic differences between language learners and proficient users.

EFL learners in this study refer to the non-English majors studying at a large university in China. The students were in their second year at the time of data collection. As part of a practical writing course requirement, the students were required to write a personal statement. The students were given two-hour in-class instruction about the basics of personal statements such as what it is, what to include, and pitfalls in writing a personal statement. The students were also given personal statement examples as a reference. With these preparations, the students were required to write their own personal statements, ranging from 600 to 800 words, outside of class. When the students finished their writing, the researchers collected them. Two students' writings were excluded because of their particular syntactic structures (they used parallel sentence structures for the whole texts). All in all, 38 of these texts were collected.

Personal statements written by English proficient users were also collected. These came from sample personal statements posted on university websites in both Canada and the United States and were chosen first because they had been posted by these universities as good examples of personal statements and second because they were highly accessible from the Internet. One of the selected personal statements exceeded 1000 words, and because the syntactic analysis software used could only analyze essays no longer than 1000 words, it was cut

short. The programs that those applicants applied to were not considered as a factor. Since the focus was on syntactic complexity, the researchers assumed that the program and overall length of the writing would not be a factor. A total of 15 personal statements by proficient users were collected and all the data were filed and made into text files for analysis. The data information is summarized in the following table.

Table 1. Data information

English writing PS	Text number	Average words	Words total	Data source
EFL learners	38	620.6	23582	Course writing
English proficient users	15	748.2	11223	American or Canadian University websites

3.2 Data Analysis Tool and Measures

The data were put into the Syntactic Complexity Analyzer to test the syntactic complexity indexes of EFL learners' writing and those of the proficient users. The Syntactic Complexity Analyzer was developed by Dr. Lu Xiaofei at Pennsylvania State University in 2010 and is open to public use by accessing it at <http://www.personal.psu.edu/xx113/downloads/l2sca.html>. The software analyzes the data using Stanford Parser and also Treegex. After reviewing the literature on syntactic complexity, Lu (2010) put forward 14 syntactic complexity indicators, including length and density measurement, for a holistic assessment of the syntactic complexity development of language learners. After the syntactic indexes statistics were generated, the statistical differences between these two groups were compared through SPSS, using independent T test.

The fourteen indicators adopted in this study (Lu, 2010) were classified into several groups. The first group concerns the length of production units. There are three indicators in this group: mean length of sentence, mean length of T unit, and mean length of clause. The second group focuses on the internal structures and is further divided into three subcategories: subordinating structures, coordinating structures, and coordinate phrases per clause. The third group is called particular structures; these include verb phrase and complex nominal structures as measurements. The specific indexes are discussed in the following results and discussion section.

4. Results and Discussion

4.1 Results of Syntactic Length Units

The length units in syntactic complexity measurement include mean length of sentence (MLS), mean length of T unit (MLT), and mean length of clause (MLC).

Table 2. Length comparison

	EFL learners	English proficient users	Independent T-test P value
MLS	22.228	26.944	0.030*
MLT	19.333	22.596	0.068
MLC	10.343	12.338	0.008*

Note. *indicates that the differences between these two groups have statistical significance.

Among the three indicators to measure syntactic lengths, the average length of sentences and clauses produced by EFL Chinese students is much lower compared to those of the English proficient users, and the differences have statistical significance as indicated by the independent T-test, with a P value of 0.03 and 0.008 respectively. There are also differences between the average length of T-unit, but with a P value of 0.068 (>0.05), there was no statistical significance.

In the length measures, Wolfe-Quintero et al. (1998) holds that the mean length of T-unit (MLT) and mean length of clause (MLC) are able to determine syntactic development in L2 writing. In this study, the distinction between these two groups measured by MLT is not as good as when measuring by mean length of sentence (MLS) and mean length of clause (MLC). Lu (2011) argued that the best length measure to distinguish L2 writing proficiency is MLC, the second being MLS, and the third being MLT. The data from the current study shows that the MLC of Chinese students is the index that most distinguished them from the proficient users and therefore this result is consistent with the results of Lu (2011). The second difference between EFL Chinese students and the proficient users is MLS. The third is the differences in MLT. With such results, this study is consistent with two other papers that studied Chinese students' syntactic complexity (Bao, 2009; Xu, 2013). That is, in terms of length indexes, the more proficient users tend to produce longer sentences and longer clauses. However, the

differences in MLT in this study failed to show any statistical significance.

4.2 Results of Subordinating or Coordinating Measurement

In the measurement of syntactic complexity, Lu (2010, 2011) classified the other eight measures into three groups. The first group belongs to sentence complexity measure, including the number of clauses per sentence (C/S). The second group concerns subordinating structures, including clauses per T-unit (C/T), dependent clauses per clause (DC/C), dependent clauses per T-unit (DC/T) and complex T-unit ratio (CT/T). The third group addresses coordinating structures, including T-units per sentence (T/S), coordinate phrases per T-unit (CP/T), and coordinate phrases per clause (CP/C).

Table 3. Subordinate or coordinate syntactic complexity comparison

	EFL learners	English proficient users	Independent T-test P value
C/S	2.204	2.206	0.987
C/T	1.894	1.847	0.617
DC/C	0.364	0.338	0.338
DC/T	0.709	0.645	0.412
CT/T	0.488	0.488	0.995
T/S	1.155	1.194	0.242
CP/T	0.606	0.797	0.190
CP/C	0.327	0.447	0.166

In terms of the density of subordinate or coordinating structures, the L2 texts written by EFL Chinese students are different from the texts written by proficient users to various degrees in these eight measures. More specifically, the EFL Chinese use more in terms of the number of DC/C and DC/T. In coordinating structures, the EFL students use fewer in terms of the number of CP/T and CP/C in comparison with their proficient counterparts. Yet none of these measure differences showed statistical significance.

From the data in Table 3, it could be inferred that the EFL Chinese students in the study, in comparison to their proficient counterparts, used more dependent clauses and fewer coordinating structures in their sentence structures. Two previous papers that were focused on EFL Chinese students came to different conclusions on this. While Bao (2009) concluded that C/T and DC/C did not distinguish language proficiency in L2 writing, Xu (2013) found that C/T in general is following a linear development from lower to higher for the EFL Chinese students; therefore Xu's paper supported the findings of Wolfe-Quintero et al. (1998 p. 85). That is, with the increase in their language proficiency, the L2 users tended to produce higher numbers of clauses in their T units. In this study, the C/T and DC/C of EFL Chinese students were not significantly different from those of proficient users. If the students in this study were evaluated as intermediate in terms of language proficiency and their counterpart as proficient in language use, then it follows that the proficient users would produce higher numbers of C/T and DC/C. However, this study failed to produce any significant differences. Therefore, this study supports the conclusion of Bao (2009) and is inconsistent with the findings of Xu (2013).

4.3 Particular Structures Measure Results

Besides the above-mentioned 11 length and clause level complexity measurements, there are three other measures that are classified as particular structures. These include verb phrases per T-unit, (VP/T), complex nominals per T-unit (CN/T) and complex nominals per clause (CN/C).

Table 4. The particular structure comparison.

	EFL learners	English proficient users	Independent T-test P value
VP/T	2.538	2.551	0.926
CN/T	2.171	2.869	0.001*
CN/C	1.168	1.561	0.001*

Note. *indicates that the differences between these two groups have a statistical significance.

Among the three particular structures, the number of verbal phrases per T unit by EFL Chinese students is similar to that of their proficient counterparts. McNamara, Crossley, & McCarthy (2010) found that the complexity of verb phrases could indicate writing quality but there is no significant difference between the two

groups here. However, there are quite large differences between the numbers of complexity nominals, particularly the number of complex nouns in clauses. According to the definition in Lu (2011), complex nominals include (1) nouns plus adjective, possessive, prepositional phrase, adjective clause, participle, or appositive; (2) nominal clauses; and (3) gerunds and infinitives in subject, but not object position.

In terms of complex nominals, the EFL Chinese students use much fewer compared to proficient users, and the difference has statistical significance. Lu (2011) pointed out that there are two best measures to predict L2 writing proficiency. One is the mean length of clause (MLC), the second being complex nominal (CN) structures. This means that the complexity measures would not be confined to the T-unit, which is consistent with some recent studies such as Biber, Gray, and Poonpon (2013), who argued that complexity at phrasal level plays a more important role in writing quality. Syntactic complexity, as a way to measure linguistic development or language proficiency, needs to reflect the related indexes of language development in a balanced way. The complex nominal structures put forward by Lu (2011) include noun phrases, adjective phrases as well as noun clauses and attributive clauses. Therefore, it is still yet impossible to conclude whether the phrase structure in the complex nominal structure would be an effective measure over the traditional measure using clause or T unit. Future studies should focus on how independent phrases affect syntactic complexity.

5. Conclusion

This study compared the syntactic differences between EFL Chinese learners and proficient users with similar writing tasks and writing time and offers several findings. First, EFL Chinese learners differ greatly from their language proficient counterparts in their use of complex nominals. Secondly, the mean sentence length and the mean clause length of EFL Chinese learners were also found to be lower compared to proficient users. Both of these tests showed statistically significant differences. Third, EFL Chinese learners were found to use more clauses and fewer coordinate structures than the proficient users, but those failed to produce statistically significant differences.

Based on such results, writing instructors could encourage EFL Chinese students to increase their sentence length or the clause length in their writings, particularly in academic writing. This could be carried out through a combination of shorter sentences. However, it should be noted that the students should not only be encouraged to increase their sentence length as the only purpose because not all proficient users employ this as the only way to indicate sophistication in their language. The more proficient users achieve this through other techniques such as the use of phrasal structures. Writing instructors could encourage students to increase their syntactic complexity through the use of phrasal structures such as noun phrases, adjective phrases, and prepositional phrases. Of course, the ultimate purpose is not to increase syntactic complexity but rather the students need to have a variety of syntactic structures at their disposal so as to improve their writing.

The study screened the data through a match between writing tasks and writing time of EFL Chinese learners and proficient users. Also while the study focused on the differences between groups, it should be acknowledged that there would likely be differences within a group. Therefore, future studies should involve more collected samples to observe the differences caused by sample size. This would also make it easier to examine differences within groups.

Acknowledgements

This paper is supported by National University Foreign Language Teaching and Research Project (#2015BJ0059B) and School of Foreign Studies, Central University of Finance and Economics. I would also like to thank my supervisor Dr. Gulbahar H. Beckett, professor in English Department, Iowa State University for her valuable suggestions and generous support.

References

- Bao, G. (2009). Syntactic complexity in EFL Learners' essays: A multidimensional perspective. *Foreign Languages Teaching and Research*, 4, 291-297.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45, 5-35. <http://dx.doi.org/10.5054/tq.2011.244483>
- Biber, D., Gray, B., & Poonpon, K. (2013). Pay Attention to the Phrasal Structures: Going Beyond T-Units-A Response to WeiWei Yang. *TESOL Quarterly*, 47, 192-201. <http://dx.doi.org/10.1002/tesq.84>
- Freeman, D. L. (1978). An ESL index of development. *TESOL Quarterly*, 12, 439-448. <http://dx.doi.org/10.2307/3586142>

- Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *The Modern Language Journal*, 80(3), 309-326. <http://dx.doi.org/10.1111/j.1540-4781.1996.tb01613.x>
- Hinkel, G. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37, 275-301. <http://dx.doi.org/10.2307/3588505>
- Hunt, K. (1965). *Grammatical Structures Written at Three Grade Levels*. National Council of Teachers of English Research report No. 3. Champaign, IL, USA: NCTE.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496. <http://dx.doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2011). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. *TESOL Quarterly*, 45, 36-62. <http://dx.doi.org/10.5054/tq.2011.240859>
- McNamara, D., Crossley, S., & McCarthy, P. (2010). Linguistic features of writing quality. *Written Communication*, 27, 57-86. <http://dx.doi.org/10.1177/0741088309351547>
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24, 492-518. <http://dx.doi.org/10.1093/applin/24.4.492>
- Qin, X., & Wen, Q. (2007). *Chinese college students' English writing development*. Beijing, Chinese Social Science Publishing.
- Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing*, 23, 497-519. <http://dx.doi.org/10.1191/0265532206lt339oa>
- Shen, J., & Bao, G. (2010). Effects of EFL proficiency and genre on the T-unit length of EFL learners' essays. *Journal of Nanjing University of Technology (social science edition)*, 4, 73-76.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27, 657-677. <http://dx.doi.org/10.2307/3587400>
- Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What Linguistic Features Are Indicative of Writing Quality? A Case of Argumentative Essays in a College Composition Program. *TESOL Quarterly*, 47, 420-430. <http://dx.doi.org/10.1002/tesq.91>
- Wolfe, Q. K., Inagaki, K., & Kim, S. H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii Press.
- Xu, X. (2013). A Study on the syntactic complexity of English Essays written by Chinese students of English. *Foreign Languages Teaching and Research*, 2, 264-275.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).