The Relationship between Syntactic Complexity and Writing Quality of Chinese EFL Learners

Jing YAN [1a*], Xiao-Yan XU [1b]

[1]School of Foreign Languages, Southwest Jiaotong University, Chengdu, Sichuan, China

[a*]144980102@qq.com; [b]470198947@qq.com

*Jing YAN

Keywords: Syntactic Complexity, Unit Length, Sentence Types, Writing Quality.

Abstract. This study analyzes the relationship between syntactic complexity and writing quality. The undergraduate participants were required to compose argumentations which were further divided into three level groups. The unit length including T- unit length and clause length, unit complexity including T-unit complexity ratio and dependent clause ratio, clauses and reduced structures are measured to find the difference across three levels and their relationship with the writing quality respectively. The result shows that unit length, unit complexity and the frequency of clauses increase with the improvement of writing quality. The frequency of adverbial clauses and object clauses discriminates middle and lower groups, while the frequency of attributive clauses discriminates higher group and middle group. The frequency of adverbial phrases is significantly different across the three groups. But there is no significant difference across the three groups in the nominal verb phrases. Except for objective clauses, appositive clauses and nominal verb phrases, other types of clauses and phrases are in the positive correlation with writing quality.

Introduction

As an important index of the first language acquisition [1], syntactic complexity, indicating the syntactic maturity, has also caught much attention in second language writing acquisition [2, 3, 4, 5]. According to different research purposes, the study of relationship between syntactic complexity and writing quality can be classified into two fields. One field studies the possible indexes for the evaluation of writing development [4, 6, 7].The agreement on the indexes has not been reached yet. For example, Larsen-Freeman concluded that T-unit length and clause length were the best indexes for syntactic maturity [8], while Gaies pointed out the limitations of T-unit as an index of syntactic complexity [9]. Wolfe-Quintero considered that clauses per T-unit and ratio of subordinate clauses to clauses were the best indexes [7]. Biber further pointed out that subordinate clauses typify the oral English while reduced structures were the feature of written English [10]. The other field studied the difference in syntactic complexity among learners of different ages, mother languages and writing competences [1, 2, 3, 5, 11].

Among the studies in China, Wen found that oral English of Chinese students manifested the characteristics of written English while their written English manifests characteristics of oral English [12]. Other studies found that their syntactic structures lacked diversity, so we should pay more attention to their syntactic complexity [13]. Another research domain is reduced phrases. There was a significant difference between Chinese English learners and native English speakers in terms of the quantity and types of reduced phrases [14]. With regard to indexes of syntactic density and length, students could write longer and longer sentences with their writing level [15, 17], but got no progress in density [15]. This study took a big sample, based on the writing development pattern proposed by Berninger [15].

Research Design

Research Questions

In what ways the higher group, the middle group and the lower group are different?

Is there correlation between writing quality and the indexes of syntactic complexity?

Participants

Sample writings of Chinese EFL learners contain timed essays. One thousand three hundred and eighty nine essays composed by students were collected from universities, such as Southwest Jiaotong University, Sichuan University, Wuhan University. The illegible writings were cast away with one thousand one hundred left for study.

Testing and Scoring

To get their writing essays, a writing test was administered to them in class period of 45 minutes. Students were asked to write a 300 words long essay discussing the topic—is Internet a wonderful invention? Full score is fifteen and the criteria is made based on the Jacob' rating scale and mark label of CET-4 and 6. Ten experienced English teachers scored the writings in way of double evaluation. Then the average score was the final score. Counting the syntactic indexes was also conducted by these ten teachers after training.

Measurement of Syntactic Complexity

The indexes of syntactic complexity were proposed by Ortega [5]. T-unit includes a main clause with its subordinate clauses. Simple sentence and subordinate compound are one T-unit, a compound sentence two. Clause is a C-unit, a subject-predicate structure. Subordinate compound has two clauses. The indexes are listed in table 1 below.

Table 1 Indexes of syntactic complexity

Measurement types	Indexes	Way of measurement
Unit length	W/T	number of words per T-unit
	W/C	number of words per C-unit
Unit density	C/T	Number of clauses/number of T-units
	DC/C	Number of subordinate clauses/number of clauses

Sentence types is another index of complexity, which are adverbial, attributive, object, subject, predicative, appositive clauses. Reduced phrases, that is, near clause phrases, are contracted from subordinate clauses. Nominal verb phrase includes gerund, infinitive phrases. Adjective phrase includes –ing participle, -ed participle. Adverb phrase includes –ing participle phrase, -ed participle phrase, -ing absolute structure, with + NP. + infinitive, with + NP+ -ing participle, with + NP + -ed participle.

Results

Length and Density

The one thousand one hundred writings were divided into three groups, that is, the higher group, the middle group and the lower group according to the writing scores with hierarchical clustering. The higher group contained 317 writings with the cluster center 11.86. The middle group contained 669 writings with the cluster center 8.87. The lower group contained 115 writings with the cluster center 5.02.

ANOVA analysis was further conducted for the three groups. There was significant difference among the three groups in terms of W/T, W/C, C/T, DC/C (p<0.001) , which meant that the higher the writing level, the longer the sentences were. Then the post hoc test, Tukey HSD, was done with the results followed.

Table 2 Post hoc test for three groups

	Level (I)	Level (J)	Mean Difference (I-J)	Std. error	Sig.
W/T	higher	middle	1.762*	.203	.000
		lower	2.492*	.291	.000
	middle	lower	.730*	.249	.012
W/C	higher	middle	.875*	.112	.000
		lower	1.056*	.179	.000
	middle	lower	.181	.166	.518
C/T	higher	middle	.052*	.016	.004
		lower	.104*	.023	.000
	middle	lower	.053*	.020	.031
DC/C	higher	middle	.029*	.007	.000
		lower	.060*	.012	.000
	middle	lower	.031*	.011	.016

*The mean difference is significant at the 0.05 level.

In terms of W/T, C/T and DC/C there is significant between-level differences which indicates that sentence length can distinguish writing levels. In terms of C-unit length, there is significant difference between higher group and middle group, but no difference between middle and lower groups, which means W/C is not as good an index as the other three to distinguish writing levels.

Sentence Types

Sentence types can be classified into that of subordinate clauses and that of reduced phrases. It is found that the use of adverbial clauses, attributive clauses and predicative clauses increases with advances of the writing level.

As to the total number of clauses (Sig.=0.000<0.05) and adverbial clauses (sig.=0.002<0.05), there is significant difference across the three groups. Significant difference could be found in the total number of object clauses (sig.=0.006<0.05) and attributive clauses (sig.=0.000<0.05), and no difference is found in the total number of subject (sig.=0.195>0.05), predicative (sig.= 0.07>0.05), and appositive clauses (sig.=0.463>0.05). Tukey HSD test was conducted for those types with significant difference across groups.

Table 3 Post hoc test for the use of subordinate clauses among three groups

	Level (I)	Level (J)	Mean Difference (I-J)	Std. error	Sig.
total number of subordinate clauses	higher	middle	.803*	.239	.002
		lower	2.261*	.382	.000
	middle	lower	1.459*	.355	.000
total number of adverbial clauses	higher	middle	.266	.113	.114
		lower	.730*	.213	.002
	middle	lower	.464*	.197	.050
total number of object clauses	higher	middle	-.124	.113	.516
		lower	.415*	.151	.017
	middle	lower	.539*	.135	.000
total number of attributive clauses	higher	middle	.495*	.134	.001
		lower	.870*	.189	.000
	middle	lower	.375	.165	.063

*the mean difference is significant at the 0.05 level.

The frequencies of adverbial and object clauses can discriminate the middle group and the lower

group. The frequency of attributive clauses can discriminate the higher group and the middle group. All the three groups rarely used nominal clauses, indicating that all the groups failed to use static language freely.

Reduced structures, which may cut short the sentences but increase embeddedness, marked the top level of syntactic development stages [7]. The use of adverbial clauses, adjective clauses and nominal clauses were studied. But the frequency of nominal clauses was an exception, with the middle group and lower group using fewer nominal verb phrases than the higher group. The ANOVA was further conducted with the result that there is significant difference among the three groups in terms of number of reduced phrases (sig. =0.000<0.05), adverbial phrases (sig.=0.000<0.05) and adjective phrases (sig.=0.000<0.05). The total number of nominal verb phrases was an exception in the sense that no difference was found across the three groups. The post hoc test is followed.

Table 4 Post hoc test for reduced phrases of three levels

	Level (I)	Level (J)	Mean Difference (I-J)	Std. error	Sig.
Total number of reduced phrases	higher	middle	.504*	.092	.000
		lower	.720*	.118	.000
	middle	lower	.225	.098	.060
Total number of adverbial phrases	higher	middle	.198*	.062	.004
		lower	.362*	.078	.000
	middle	lower	.164*	.065	.034
Total number of adjective phrases	higher	middle	.327*	.059	.000
		lower	.380*	.080	.000
	middle	lower	.053	.065	.696

As can be seen from table 4, the higher group used significantly more reduced phrases, more adverbial and adjective phrases than the other two groups, which meant that the use of reduced phrases and the adjective phrases could discriminate high group and the other two groups. There was no difference in nominal phrases across the three groups, which indicates that students have poor command of nominal verb phrases.

Correlation between Writing Quality and Each Index

In order to study the relationship between the writing quality and each index, correlation study was carried out. Significant correlations were found between writing quality and W/T (r=0.325, P<0.01), W/C (r=0.238, P<0.01), C/T (r=0.164, P<0.01), DC/C (r=0.179, P<0.01), subordinate clauses (r=0.187, P<0.01), adverbial clauses (r=0.113, P<0.01), attributive clauses (r=0.151, P<0.001), subjective clauses (r=0.073, P<0.05), predicative clauses (r=0.079, P<0.01), reduced phrases (r=0.195, P<0.01), adverbial phrases (r=0.148, P<0.01), adjective phrases (r=0.162, P<0.01). The appositive (r=0.016, P>0.05), object clauses (r=0.037, P>0.05) and the nominal verb phrases (r=-0.217, P>0.05) hold no correlations with writing quality.

Discussion

Difference among the Three Groups In Terms of Syntactic Complexity

The average difference in sentence length between the adjacent groups is about two words, which indicates that sentence length is a reliable index of writing quality. The longer the sentence length, the better the writer can command syntactic structure and vocabulary. Meanwhile, the standard deviation of high group is the largest (3.188), which indicates that higher group can use sentences of various lengths flexibly and skillfully. The results of unit length are similar to that of Bao [16]. The average sentence length got in this study is 15.072 while 14.08 in his study. The difference may lie in the data grouping.

The unit density will rise with the writing level, and high density indicates that students can

organize sentences logically by using clauses of various kinds of logical relations to make the meaning closely connected. However, in Bao's study [16], there was no significant difference in unit density among groups.

In terms of syntactic diversity the most frequently used types are adverbial clauses, object clauses, and attributive clauses while the other sentence types are seldom used, which accords with the findings of Qin and Wen [13]. The use of adverbial and object clauses can discriminate the middle and lower groups. The lower group is not good at organizing information logically by using less clauses and more simple sentences with bamboo sentence patterns. The middle group is conscious of using clauses, but only basic and simple types. All the three groups seldom use subject, predicative and appositive clauses, and there is no difference among them, which implies that students in general have a poor command on nominal clauses. This is in accordance with the scales of syntactic complexity development proposed by Biber on the basis of corpus [4]. According to the scale tablet by Biber, the lowest level is object clauses, then the adverbial clauses, which are the first two types of clauses acquired firstly by students [4]. What should be marked is that the use of adverbial clauses concentrates on those led by *when, because, if*, not those led by *as, while*, which is similar to the clauses distribution of the lower group in Myhill's study [1]. According to Biber, the attributive clause is placed on the third level and in this study it can discriminate the higher group and the middle group [4]. Although attributive clause has gained much attention in language teaching, Chinese students fail to grasp it well because of negative transfer of mother language in which the attributive clause is positioned before the main clause. Predicative and subject clauses are also placed on the third level, but Chinese learners fail to command them well as compared to attributive clause because less attention is given to them than attributive clause. Appositive clause is placed on the fifth level which is the highest level of syntactic complexity and the last clause type which is acquired by students. Three groups seldom use it with no significant difference.

Reduced structures are the important sign of written language. In the study of syntactic complexity, complexity is defined as using nominal verb structure, simplifying the clauses to prepositional phrases, adding subordinate structures or changing the types of subordinate structures [4]. Reduced structures indicate a higher level of syntactic ability, which can also discriminate the higher group and the middle group. In the syntactic acquisition order, nominal verb structure is the most difficult, then the adjective phrases, and then the adverbial phrases. The skill to transform subordinate clauses to participles, gerunds and infinitives should be trained massively, otherwise learners would avoid using the non-finite structures. Besides, good command of adverbial and verbal phrases means good command of adverbial clauses and attributive clauses.

Relationship between Indexes and Writing Quality

Most of indexes of syntactic complexity are positively correlated with writing quality. As compared with the narrative writing, argumentative writing requires more syntactic resources, so high syntactic complexity may help the progression of writing quality.

In terms of length, W/T, which manifests syntactic complexity on T-unit, and writing quality are highly correlated W/T (.325**). It can reflex to some degree learners' use of subordinate clauses and other reduced structures, so it can be an indicator of writing quality, which is in agreement with the finding by Qin and Wen [13]. According to Hunt's study, the best indicator for the syntactic maturity is T-unit length which can be raised by complex subordinate structures, embedded structures. W/C (.238*) shows complexity on level of clauses which are loaded by rich information with condensed structures. So C-unit length can also reflect the writing quality. In Lu's research [17], the result showed that C-unit length was the best indicator of language proficiency. In this study, the correlation between writing quality and W/C is lower than that between writing quality and W/T, and this can be explained by the fact that Chinese learners seldom use reduced structures. In Xu's study, the clause length was significantly shorter than that of native speakers, and the quantity of reduced structures, especially the adjective phrases, were significantly less than that of the native speakers [14]. Because the learners use few reduced clauses, the length of the clauses becomes short accordingly, which in turn affects the correlation between clause length and writing

quality.

In terms of density indexes, C/T (.164**) and DC/C (.179**), significantly correlated with writing quality, measure clauses with finite verbs or subordinate clauses, which manifest ratio of embedded clauses. The higher the ratio of embeddedness, the more clauses in T-unit or C-unit and the better the writing quality.

In syntactic diversity, the total number of clauses (.187**), subordinate clauses, except for appositive clauses, are significantly correlated with writing quality. The higher the writing quality, the more diverse the clauses. Learners on every level seldom use appositive clauses which explains the head noun, reflecting the rigidity of the writer. No correlation is found between writing quality and appositive clause, probably because learners might be affected by negative transfer of Chinese which is a left-branching language without post-modifiers.

The number of reduced phrases is significantly correlated with writing quality, higher than subordinate clauses except attributive clause. Various types of reduced phrases can increase syntactic complexity and diversity. The use of reduced phrases requires higher syntactic proficiency, because all types of the phrases are placed at the highest level of complexity. Reduced phrase is the indicator of written language. By using all types of reduced clauses, learners can change colloquial clauses into more concise phrases.

It should be noted that the correlation coefficients are not so high, because writing quality is determined by various factors including vocabulary, discourse organization, and content [18]. Syntactic complexity is the necessary but not sufficient condition for high quality writing.

Conclusion

This study focused on the relationship between syntactic complexity and diversity and writing quality. Generally, unit length, unit density, use of clauses and reduced phrases rise with the betterment of writing. Adverbial clauses and object clauses can discriminate the middle and the lower groups, while attributive clauses can discriminate the middle and the higher groups. With regard to the reduced structures, there is significant difference among the three groups in terms of adverbial phrases. Adjective phrases can discriminate the middle and the lower groups. No difference is found among the three groups in terms of nominal verb phrases.

According to correlation analysis, indexes of complexity and diversity, except object clause, appositive clause, and nominal phrase, hold positive correlation with writing quality. The order of correlation is W/T > W/C> reduced structures > number of clauses > DC/C > C/T > adjective phrases> attributive clauses > adverbial phrases > adverbial clauses >predicative clauses > subject clauses.

References

[1] Myhill, D. 2008, Towards a linguistic model of sentence development in writing. Language and Educati*on* 22: 271-288.

[2] Philip, S. and Eric T. L. 1998, What develops in the development of second language writing? *Applied linguistics* 19: 224-254.

[3] Larsen-freeman, D. 2006, The emergence of complexity, fluency and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics* 27/4 590-619.

[4] Biber, D., Gray, B. & Poonpon, K. 2011, Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly 45:*15-35.

[5] Ortega, L. 2003, Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied linguistics* 24:429-518.

[6] Gaies, S. J. 1980 T-Unit analysis in second language research: application, problems and limitations. *TESOL Quarterly* 14: 153-59.

[7] Wolfe-Quintero, K., Inagaki, S., & Kim, H. _Y. 1998. Second language development in writing: measures of fluency, accuracy and complexity. Honolulu: University of Hawaii Press.

[8] Larsen-Freeman, D. 1978, An ESL index of development. *TESOL Quarterly* 12: 4, 439-448.

[9] Gaies, S. J. 1980, T-Unit analysis in second language research: applications, problems and limitations. *TESOL Quarterly* 12: 453-60.

[10] 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: 1, 192-201.

[11] Bardovi-Harlig, K., Theodora, B. 1989, Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition,* 11: 17-34.

[12] Wen, Q. F. 2009. A study on changes of register features in learner English. *Journal of Foreign Languages*, 32: 2-10.

[13] Qin, X. Q., Wen, Q. F. 2007. EFL writing of college English majors in China. Beijing: China Social Sciences Publishing House.

[14] Xu, X. Y., Sun, N. H. 2012. Introducing cognitive senses into the teaching and learning of English reduced structures. *Journal of Southwest Jiaotong University (Social Science)*, 13:20-26.

[15] Berninger, V. W., Fuller, F. and Whittaker, D., 1996, A process model of writing development across the lifespan. Educational Psychology Review 8 (3), 193-217

[16] Bao, G. 2009. Syntactic complexity in EFL learners' essays. Foreign Language Teaching and Research, 41: 4, 291-297.

[17] Xiaofei, L. 2011, A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers language development. TESOL Quarterly, Vol.45, No. 1, 36-62

[18] Ma, G. H., Wen, Q. F. 1999.The relationship of L2 learners' linguistic variables to L2 writing ability. Foreign Language Teaching and Research, 4: 34-39.