



# The Effect of Online Writing on the Syntactic Complexity of EFL Learners' English Compositions

Gu Tongqing

School of Foreign Languages, Chengdu University of Information Technology, Chengdu, Sichuan Province, PRC, China

## Abstract

This paper explores the effect of online writing on the syntactic complexity of EFL learners' English writing among graduate students. We sampled the initial and final version of 55 English argumentative essays written by Chinese university students on an online writing platform. These essays were analyzed using 14 syntactic complexity measures with the L2 Syntactic Complexity Analyzer (Lu, 2010). Results show that there are significant differences in the use of syntactic structures between the two versions, while there is no significant differences in the 14 measures of syntactic complexity between the initial and final version. The implications of the results for L2 writing teaching and for automatic scoring system are also discussed.

**Keywords:** Syntactic structure; Syntactic complexity; Online writing.

## 1. Introduction

In second or foreign language education, computer-assisted language learning (CALL) has recently been proposed as an alternative to traditional teaching methods and techniques. According to previous studies, CALL can provide a natural means of meaningful communication, gives the L2 learner opportunities to practice language skills, for both linguistic and communicative purposes. As modern education technology is developing rapidly and teaching conditions are improving continuously, network-based CALL becomes more and more popular in Chinese schools at all levels.

As a complex process of knowledge construction, English writing has been the focus of many researchers. Some educators and researchers have discussed the characteristics and advantages of computer-aided English writing teaching, and how to introduce Internet-based computer-assisted method to the process of writing teaching. Among the 3 measures of writing quality, fluency, accuracy and complexity, both lexical and syntactic complexity are considered to be important indicators of L2 writing development. The importance of syntactic complexity, commonly construed as the variety and degree of sophistication of the syntactic structures deployed in written production (Lu, 2011; Ortega, 2003), has long been recognized, as evidenced in the large number of studies that have examined the relationship of syntactic complexity in L2 writing to L2 proficiency (Ai and Lu, 2013; Lu, 2011; Ortega, 2000;2003) or the quality of L2 writing, over the past two decades.

The current study is to examine differences in the complexity of syntactic structures in the initial and final English writing version written online by non-English major graduate students. The first and final version of students' online writing were compared to reveal the differences in syntactic complexity. Our primary focus here is to find out whether repeated revision can actually enhance syntactic complexity in L2 writing. This investigation will then be discussed in light of previous claims about the relationship between revision and L2 writing quality. Implications of our results for L2 writing pedagogy will also be discussed.

## 2. Literature Review

### 2.1. Syntactic Complexity and Second Language Writing

Syntactic complexity has been commonly characterized as the range of syntactic structures that are produced and the degree of sophistication of those structures (Ortega, 2003). Previous studies focus on the relationship of syntactic complexity in L2 writing to L2 proficiency (Ai and Lu, 2013; Lu, 2011; Ortega, 2000;2003) or the quality of L2 writing (Lu and Ai, 2015). Results from such studies have shown that some measures of syntactic complexity may be reliably used to differentiate levels of L2 proficiency, and some to predict the quality of L2 writing (Ai & Lu, 2013). Meanwhile, researchers have also found that syntactic complexity in L2 writing may be affected by various learner-, task-, and context-related factors, such as topic, genre, planning time, and instructional setting, among others (Lu, 2011; Ortega, 2003).

### 2.2. Measures of Syntactic Complexity in L2 Writing

As a multidimensional construct, there are various measures of syntactic complexity. In this study, syntactic complexity was measured by 14 indices provided in the L2 Syntactic Complexity Analyzer (Lu, 2010), the computational tool used to analyze the syntactic complexity of the writing samples. Each of these measures gauges one of the following five dimensions of syntactic complexity: length of production unit, amount of subordination, amount of coordination, degree of phrasal sophistication, and overall sentence complexity (Ai and Lu, 2013; Lu,

2010;2011). The measures, their definitions, and the dimensions of syntactic complexity they represent are summarized in Table 1, adapted from Lu (2010).

Table-1. Syntactic complexity measures

Measure	code	definition
1. Length of production unit		
Mean length of clause	MLC	# of words/# of clauses
Mean length of sentence	MLS	# of words/# of sentences
Mean length of T-unit	MLT	# of words/# of T-units
2. Amount of subordination		
Clauses per T-unit	C/T	# of clauses/# of T-unit
Complex T-units per T-unit	CT/T	# of complex T-units/# of T-units
Dependent clauses per clause	DC/C	# of dependent clauses/# of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses/# of T-units
3. Amount of coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases/# of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases/# of T-units
T-units per sentence	T/S	# of T-units/# of sentences
4. Degree of phrasal sophistication		
Complex nominals per clause	CN/C	# of complex nominals/# of clauses
Complex nominals per T-unit	CN/T	# of complex nominals/# of T-units
Verb phrases per T-unit	VP/T	# of verb phrases/# of T-units
5. Overall sentence complexity		
Clauses per sentence	C/S	# of clauses/# of sentences

### 3. Research Questions

The current study aims to answer the following research questions:

Are there significant differences in the syntactic structures used in the first and the last version of graduate students' online English writing? and, if yes, what are these differences?

Are there significant differences in the syntactic complexity of the first and the last version of online English writing of graduate students and, if yes, what are these differences?

The differences between the two versions are hypothesized to reveal the effect of online revision and the effect of online writing on students' writing performance.

## 4. Method

### 4.1. Research Data

The data used in the current study were collected from graduates students' online writing samples on the topic "Solution to the problem of plastic pollution: legislation or technology?", a big writing event on the same topic in 2019. The writing task was launched on an online writing platform ([www.pigai.org](http://www.pigai.org)), where students can register for free use. Students wrote and submitted online and automatically received score and detailed correction feedback of their writing within seconds. They can then revise their writing online according to the feedback for as many times as they like to improve the writing quality. The current study downloaded the initial and the final version of 113 graduate students' writing samples from the online writing platform, and according to the times of revision, retained only the writing samples with more than 6 revisions. Then the initial and final samples of the remaining 55 samples were saved as txt files for further analysis.

### 4.2. Data Analysis

Each essay in the final data set was analyzed using the L2 Syntactic Complexity Analyzer (Lu, 2010), a computer program designed to analyze the syntactic complexity of English writing samples using the 14 measures discussed in section 2.2. L2SCA was chosen because of its free availability, its integration of a large set of measures that are viable candidates for syntactic complexity research, its capability to process files in batches, and its high reliability (Lu and Ai, 2015). For each writing sample, L2SCA produces frequency counts for the following nine structural units: words, sentences, verb phrases, clauses, dependent clauses, T-units, complex T-units, coordinate phrases, and complex nominals; it also returns 14 indices of syntactic complexity calculated using the frequency counts. Lu (2010), reported accuracy ranging from .830 to 1.000 for structural unit identification and correlations ranging from .834 to 1.000 between the syntactic complexity scores computed by human annotators and L2SCA. After the syntactic complexity indices have been obtained for each essay in the data set, a set of paired samples t-tests were run to compare differences between the different versions of students' writing for each of the 14 syntactic complexity measures.

## 5. Results and Discussion

### 5.1. Differences Between the Score of the First and Final Writing Version

Table 2 describes the means and standard deviations of the scores as well as the significant differences between scores of the first and the final versions. The results show that the final version had significantly higher score than the initial one ( $p < .001$ ), which indicates that, through multiple revisions, students' writing quality has been improved significantly. As the online writing platform offers detailed correction feedback and suggestions on the choice of words, grammar, logical connection of the writing, students can either correct the mistakes in their writing or improve the use of words and sentences and make a new submission. Through multiple revisions, the score can be greatly improved.

Table-2. Mean values and standard deviations of the scores of two versions

	Mean	Std.	Std. error	t	df	Sig.
Compare Score	-8.56667	4.96667	.90679	-9.447	29	.000

### 5.2. Syntactic Complexity of the two Versions

One of the measures of writing quality is complexity, and both lexical and syntactic complexity are indicators of writing development. The following sections mainly describe and analyze the syntactic complexity of the two versions to see whether there is any difference between them.

#### 5.2.1. Syntactic Structures

Table 3 summarizes the mean values and standard deviations of syntactic structures and the significant differences between the two versions. The results show that there are significant differences between number of words, sentences, clauses, and T-units. The final version has significantly more words thus longer than the first version ( $p < .005$ ), more sentences, clauses and T-units than the first version, even though the differences are not so significant as the number of words ( $p > .005$ ). This is easy to explain by the fact that each revision naturally leads to the addition of words, clauses, and sentences to make the writing more convincing and complete although there might also be deletion or change of the content. A comparison of the text of the two versions shows that when making revisions, students mainly focus on the lexical level rather than syntactic and textual level revision.

Table-3. Mean values and standard deviations of syntactic structures

		mean	Std.	Std. error	t	df	Sig.
Pair 1	W1 - W2	-34.80000	63.16830	11.53290	-3.017	29	.005
Pair 2	S1 - S2	-1.90000	4.67827	.85413	-2.224	29	.034
Pair 3	C1 - C2	-2.83333	5.93112	1.08287	-2.617	29	.014
Pair 4	T1 - T2	-2.40000	5.14346	.93906	-2.556	29	.016

Note: W1=word count in initial version, W2=word count in final version; S1=sentence in initial version, S2=sentence in final version; C1=clause in initial version, C2=cause in final version; T1= T-unit in initial version, T2= T-unit in final version.

#### 5.2.2. Length of Production Unit

Table 4 summarizes the mean values and standard deviations of the three length measures as well as the significant differences between the initial version and the final in these measures. The results show that there is no significant differences between the two versions in all three measures ( $p > .005$ ). These results are in contrast to those reported in the first part of the analysis where there is significant difference in syntactic structures between the two versions. It is found that although there are more sentences, clauses, and T-units in the final version, the mean length of these syntactic units are not quite different between the two versions.

Table-4. Mean values and standard deviations of MLS, MLC, and MLT of the first and final versions

		mean	Std.	Std. error	t	df	Sig.
Pair 1	MLS1 - MLS2	.05800	4.05037	.73949	.078	29	.938
Pair 2	MLT1 - MLT2	.30700	2.95760	.53998	.569	29	.574
Pair 3	MLC1 - MLC2	-.06600	.91083	.16629	-.397	29	.694

Note: MLS1, MLT1, MLC1=mean of sentence, T-unit, clause of initial version; MLS2, MLT2, MLC2=mean of sentence, T-unit, clause of final versions

#### 5.2.3. Amount of Subordination

Table 5 summarizes the means and standard deviations of the four subordination measures and the significant differences between the two versions in these measures. These results show there is no significant differences between the two versions in all the subordination measures.

**Table-5.** Mean values and standard deviations of subordination measures of two versions

		Mean	Std.	Std. error	t	df	Sig.
pair 1	C/T1 - C/T2	.03436	.50933	.09299	.369	29	.714
Pair 2	DC/C1 - DC/C2	.01051	.19163	.03499	.300	29	.766
pair 3	DC/T1 - DC/T2	.01980	.44330	.08093	.245	29	.808
pair 4	CT/T1 - CT/T2	.03178	.29486	.05383	.590	29	.560

**Note:** The numeral “1” following C/T, DC/C DC/T, CT/T refers to the initial version, the numeral “2” refers to final version

### 5.2.4. Amount of Coordination

Table 6 summarizes the means and standard deviations of the three coordination measures as well as significant differences between the two versions in these measures. The results show there is no significant differences between the two versions in all the coordination measures. Similar to subordination measures

**Table-6.** Mean values and standard deviations of coordination measures

		Mean	Std.	Std.error	t	df	Sig.
pair 1	CP/C1 - CP/C2	.00684	.34698	.06335	.108	29	.915
pair 2	CP/T1 - CP/T2	.02407	.38147	.06965	.346	29	.732
pair 3	T/S1 - T/S2	-.01913	.20875	.03811	-.502	29	.619

**Note:** The numeral “1” following CP/C, CP/T, T/S refers to the initial version, the numeral “2” refers to final version

### 5.2.5. Degree of Phrasal Sophistication

Table 7 summarizes the means and standard deviations of the phrasal complexity measures as well as significant differences between the two versions. As shown in the table, there is no significant differences between the two versions in all these measures.

**Table-7.** Mean values and standard deviations of phrasal complexity measures

		Mean	Std.	Std. error	t	df	Sig.
pair 1	CN/C1 - CN/C2	.00424	.67668	.12354	.034	29	.973
pair 2	CN/T1 - CN/T2	.07697	.93606	.17090	.450	29	.656
pair 3	VP/T1 - VP/T2	.02791	.74668	.13632	.205	29	.839

**Note:** The numeral “1” following CN/C, CN/T, VP/T refers to the initial version, the numeral “2” refers to final version

### 5.2.6. Overall Sentence Complexity

Table 8 summarizes the means and standard deviations of the overall sentence complexity measures as well as significant differences between the two versions. The results showed there is no significant difference in overall syntactic complexity. This result shows that there is little syntactic difference between the two versions, and when making revisions, students make few syntactic changes

**Table-8.** Mean values and standard deviations of overall sentence complexity measure

		Mean	Std.	Std. error	t	df	Sig.
pair 1	C/S1 - C/S2	.01548	.61382	.11207	.138	29	.891

**Note:** C/S1 refers to the initial version, C/S2 refers to final version

The results show that even though students online writing differ significantly in syntactic structures between the 2 versions, there is little difference between the two versions in syntactic complexity. This can be explained from the following aspects. On the part the students, their’ knowledge of syntax and grammar is inadequate. As a result, they are not very confident in making sentences when making revisions so they may deliberately avoid changing original structures to avoid making new mistakes. For another, students are not aware of the importance of syntactic complexity in writing, so they may choose to use short and simple sentences more frequently because these are safe from mistakes. On the part of teacher and English grammar teaching, with the new trend of content-based language teaching, English teaching now at the university level focus more on the textual organization than on lexical and syntactic knowledge so that students pay little attention on the importance of vocabulary diversity and syntactic complexity. In addition, there is no consistent connection between syntactic complexity and writing quality as writing with high quality is not necessarily syntactically complex as mentioned by Beers (2007).

## 6. Conclusion

This study revealed there is significant differences in the number of syntactic structures, but there is very little difference in multiple dimensions of syntactic complexity in graduate students’ online English writing, although further research is needed to establish any causal links between writing score and syntactic complexity patterns in L2 writing. These results call for the need to take learners’ syntactic development into account in assessing their L2 proficiency. In addition to the implications for L2 writing research, our findings have useful implications for online writing platforms as well. Our results show that when students’ writing scores are improved, there is actually little change in terms of syntactic complexity development, even after constant revisions. It may be useful for writing

teachers to be aware of the fact that the automatic scoring may be problematic in some cases, and teachers cannot solely rely on the automatic online scoring to evaluate learners' writing performance.

## References

- Ai, H. and Lu, X. (2013). *A corpus-based comparison of syntactic complexity in NNS and NS university students writing*. In A. Díaz-Negrillo, N. Ballier, and P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data*. John Benjamins: Amsterdam.
- Beers, N. W. S. (2007). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? : Available: <http://www.springerlink.com/content/326811575526h8t8/full-text.pdf>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4): 474–96.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1): 36-62.
- Lu, X. and Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29: 16-27. Available: <http://dx.doi.org/10.1016/j.jslw.2015.06.003>
- Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners*. Unpublished doctoral dissertation. Honolulu: University of Hawaii.
- 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.