Exploring Topic Effect on Syntactic Complexity of English Writings

Zhou Bi[1]

[1]Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Zhou Bi, Guangdong University of Foreign Studies, No.2 Baiyun Avenue, Guangzhou City, China. Tel: 86-180-775-75652. E-mail: 20170110068@gdufs.edu.cn

Abstract

This paper aims to explore topic effect on syntactic complexity in Chinese EFL learners' and English native speakers' writings. To this end, the paper examined 600 argumentative essays on two topics, using 14 measures of syntactic complexity to investigate whether there is significant difference in topic effect on EFL writings and ENS writings. Four dimensions are included in these 14 syntactic complexity measures, which are *length of production unit*, *amount of subordination*, *amount of coordination* and *degree of phrasal sophistication*. Results showed strong topic effect on both Chinese EFL writers and ENS writers in terms of *length of production unit* and *particular structure*. The topic effect differs in *amount of coordination* in that the Chinese EFL group produced different amount of coordination in the two topics, while the ENS group did not show significant difference in this aspect.

Keywords: syntactic complexity, language teaching, English writings, topic effect

1. Introduction

The issue of syntactic complexity has received considerable attention from researchers in a wide range of fields. In second language research, different measures of syntactic complexity have been proved to be effective in predicting writing quality (Crowhurst, 1980; Beers and Nagy, 2009; Yang, Lu, & Weigle, 2015), but to a different degree. Previous studies have examined the effect of L2 proficiency (Lu, 2011; Ai & Lu, 2013), L1 background (Lu & Ai, 2015) task-related factors such as genres, timing condition, topics (Yang, Lu & Weigle, 2015; Biber et al, 2016; Yoon & Polio, 2017; Yoon, 2017), on syntactic complexity.

In areas of L2 writings and language assessment, there has been increasing attention to the task-related factors such as genres, task types and topics. Several attempts have been made to investigate the **genre difference**. For example, Lu (2011) examined the linguistic complexity in argumentative and narrative essays from Chinese students (EFL) and found greater linguistic complexity in the argumentative essays (12 of 14 syntactic measures). Yoon and Polio (2017) did a conceptual-replicated study and further examined the interaction effect of time and genres difference on linguistic complexity (involving syntactic measures, lexical measures, fluency and accuracy). They found significant genre effect on phrasal level but not on clauses level, but no genre effect on fluency and accuracy. The results of their study resonate with those of Lu (2011)'s study in some way. But the study only lasted for one semester, which might not be enough to shed light on the developmental process of linguistic complexity in L2 writings.

1.1 Previous Studies of Topic Effect on Syntactic Complexity

The studies of genres and task types have produced somewhat consistent findings. But regarding to the topic effect, there seems to be relatively little research on the relationship between **topics** and syntactic complexity. Yang, Lu and Weigle (2015) examined the role of topic in L2 writings and found that topic had a significant effect on syntactic complexity. It reported that the *future* topic which demands more causal reasoning tends to elicit a significantly higher amount of subordination while the *appearance* topic with less causal reasoning demand elicited a higher amount of elaboration at the finite clause level. However, this was challenged by Yoon (2017) that, given their diverse background (38 L1 backgrounds) the participants might not perceive the two topics as intended by the researchers, in terms of the task's inherent level of reasoning demand. Therefore, Yoon explored the topic effect in terms of their relevance to writers' own experience. His study showed that the topic which believed to be more relevant to college-level EFL writers tends to elicit more complex language.

2. Methodology

2.1 Research Questions

As the previous studies reported, there are significant topic effects on syntactic complexity of essays. However, it is still unclear whether the topic effect on syntactic complexity of learners' writing and English native speakers' writings is the same. Therefore, this study seeks to extend Yoon's research by examining topic effect on both EFL writers and English native speaker writers and it is hoped to find out if similar topic effect is unique to L2 writers. It seeks to answer the following questions:

1) How do EFL and ENS writings differ in syntactic complexity in terms of the same topic?

2) How do EFL and ENS participants' writings differ in the measures of syntactic complexity across different topics?

To answer these questions, the study will analyse a corpus of argumentative essays on two topics written by Chinese EFL learners and English native speakers. The main goal of this study is to find out the different effect of the two topics on syntactic complexity of English writings, and by comparing the results of EFL learner group and ENS group, to find out if topic effects differ across EFL and ENS writers. Therefore, this is a 2 ×2 design.

2.2 Corpus Data

The corpus adopted in this study is the *International Corpus Network of Asian Learners of English (ICNALE)* (Ishikawa 2013). The written module of the corpus includes argumentative essays written by Asian English learners as well as English native speakers. The ICNALE rigidly controls prompt, tasks, other parameters such as time for writing, the length of an essay, which helps to guarantee the reliability of a contrastive analysis. The corpus data was last updated on 15th April, 2018.

To serve the purpose of this study, 600 essays are randomly selected as analysing data, including 300 essays on the topic of part time job (150 written by Chinses EFL learners and 150 by English native speakers) and 300 essays on the smoking topic (150 by each group). Table 1 summarizes the composition of the selected corpus data and word number of the essays. The *language background* in this study refers to EFL (English as foreign language) and ENS (English as mother tongue).

Table 1. Composition of the selected corpus

Topic	Language background	Number of essays	Numbers of words	Words per essays Mean (sd)
PTJ	CHN	150	35,024	233.5(29.4)
	ENS	150	33,942	226.3(26.4)
SMK	CHN	150	34,492	229.9(27.2)
	ENS	150	34,106	227.4(26.4)
Total		600	137,564	229.3(26.7)

Note: NNS: non-native speakers

There are only two topics, which are: (a) It is important for college students to have a part-time job (PTJ), and (b) Smoking should be completely banned at all the restaurants in the country (SMK). As suggested in Yoon (2017), the part-time job topic is more relevant to college students while the smoking banning is more concerned with public life. Therefore, the major difference between the two topics is the relevance to college students. In this regard, all the essays collected are written by college students.

2.3 Instruments

In previous studies, many researchers only examined at most four or five measures due to the intensive annotation work. However, it is necessary to examine syntactic complexity with multi-dimensional measures (Norris & Ortega, 2009). Therefore, this study will apply an automated processing tool, L2 Syntactic Complexity Analyzer online[1], developed by Lu (2010). It returns 14 indices of syntactic complexity as well as nine structural unites calculated by frequency counts. Table 2, cited from Lu (2010), shows 14 syntactic complexity measures and definitions of each measure.

Table 2. Syntactic complexity measures produced by L2SCA

Measure	Code	Definition
Type 1 Length of production unit		
Mean length of sentence	MLS	# of words/#of sentences
Mean length of T-unit	MLT	# of words/#of T-units
Mean length of clause	MLC	# of words/#of clauses
Type 2 Sentence complexity		
Clauses per sentence	C/S	# of clauses/#of sentences
Type 3 Amount of subordination		
Clauses per T-unit	C/T	# of clauses/ # of T-units
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
Type 4 Amount of coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases/ # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrase/ # of T-units
T-units per sentence	T/S	# of T-units/ # of sentences
Type 5 Particular structures		
Complex nominals per clause	CN/C	# of complex nominals/ # of clauses
Complex nominals per T-units	CN/T	# of complex nominals/ # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases/ # of T-units

2.4 Analytical Procedure

This study attempts to find out the effect of two topics (part-time job; Smoking banning) on syntactic complexity in writings. The basic assumption is that the topic effects are different in EFL writings (hereafter CHN) and ENS writings. Therefore, this study will first analyze the fourteen measures of syntactic complexity of all 600 essays (by using only LASCA). A non-parametric test will be applied to see if there is significant difference between the effect on Chinese group and English group, and an additional test for difference between the two groups given the same topic.

3. Results and Discussion

3.1 Descriptive Result

Table 3 represents the mean and standard deviation of the fourteen measures in CHN writings and ENS writings. Roughly from the table, it can be seen that in both topics, all the means of fourteen measures in ENS writings are more than those in CHN writings.

Table 3. Descriptive data of the 14 syntactic complexity measures

	CHN				ENS			
	PTJ		SMK		PTJ		SMK	
	mean	sd	mean	sd	mean	sd	mean	sd
Measure								
Length of production unit								
MLS	MLS		MLS		MLS		MLS	
MLT	MLT		MLT		MLT		MLT	
MLC	MLC		MLC		MLC		MLC	
Sentence complexity								
C/S	C/S		C/S		C/S		C/S	
Amount of subordination								
C/T	C/T		C/T		C/T		C/T	
CT/T	CT/T		CT/T		CT/T		CT/T	
DC/C	DC/C		DC/C		DC/C		DC/C	
DC/T	DC/T		DC/T		DC/T		DC/T	
Amount of coordination								
CP/C	CP/C		CP/C		CP/C		CP/C	
CP/T	CP/T		CP/T		CP/T		CP/T	
T/S	T/S		T/S		T/S		T/S	
Particular structures								
CN/C	CN/C		CN/C		CN/C		CN/C	
CN/T	CN/T		CN/T		CN/T		CN/T	
VP/T	VP/T		VP/T		VP/T		VP/T	

Note: A non-parameter test is required to see if there is significant difference between two language background or between two topics.

3.2 Language Background

Answer to RQ1: How do EFL and ENS writings differ in syntactic complexity in terms of the same topic?

Table 4 examines the difference between CHN and ENS on the same topic. Even though it only shows a rough picture over the difference between the two groups, the result is surprisingly consistent with the Lu and Ai's (2015) study comparing the writings of Non-native speakers (NNS) and Native speakers (NS). As it can be seen, there is no significant difference in mean length of clause (MLC) and complex nominals per clause (CN/C). In the PTJ topic, there is also no difference in coordinate phrases per clause (CP/C), which is also found in their study. Except for the three measures above, all the other measures differ significantly between CHN and ENS writings. Taking the means of each measure in table 3 into consideration, it is found that ENS writings have greater values in all measures of syntactic complexity. This may represent that ENS writings are generally more complex than CHN writings in both topics.

Table 4. Syntactic complexity measures comparing CHN and ENS

Measure	p value (PTJ)[a]	p value (SMK)[a]
Length of production unit		
MLS	**.000**[b]	**.000**[b]
MLT	**.000**[b]	**.000**[b]
MLC	.435	.047
Sentence complexity		
C/S	**.000**[b]	**.000**[b]
Amount of subordination		
C/T	**.000**[b]	**.000**[b]
CT/T	**.000**[b]	**.000**[b]
DC/C	**.000**[b]	**.000**[b]
DC/T	**.000**[b]	**.000**[b]
Amount of coordination		
CP/C	.101	**.000**[b]
CP/T	**.000**[b]	**.000**[b]
T/S	**.000**[b]	**.000**[b]
Particular structures		
CN/C	.261	.029
CN/T	**.000**[b]	**.000**[b]
VP/T	**.000**[b]	**.000**[b]

Notes: [a] the alpha value for the analysis was adjusted by the Bonferroni correction for multiple test. 0.05/14=0.00357

[b] $p < 0.00357$

3.3 Topic Effect

Answer to RQ2: How do EFL and ENS participants' writings differ in the measures of syntactic complexity across different topics?

Table 5 presents the result of Mann-Whitney testing respectively on the CHN and ENS group.

Table 5. Syntactic complexity measures comparing topic effect (both CHN and ENS group)

Measure	p value (CHN)[a]	p value (ENS)[a]
Length of production unit		
MLS	.079	**.001**[b]
MLT	.059	**.000**[b]
MLC	**.000**[b]	**.002**[b]
Sentence complexity		
C/S	.024	.367
Amount of subordination		
C/T	.019	.217
CT/T	.234	.136
DC/C	.140	.073
DC/T	.046	.109
Amount of coordination		
CP/C	**.000**[b]	.585
CP/T	**.000**[b]	.230
T/S	.840	.323
Particular structures		
CN/C	**.000**[b]	.009
CN/T	**.000**[b]	**.000**[b]
VP/T	**.000**[b]	**.000**[b]

Notes: [a] the alpha value for the analysis was adjusted by the Bonferroni correction for multiple test. 0.05/14=0.00357

[b] $p < 0.00357$

It can be found that in CHN group, three types of syntactic measures are significantly different across the two topics. In the first type, only mean length of clause (MLC) is shown to be significantly different. Considering the means of MLC (mean$_{PTJ-MLC}$=9.40 > mean$_{SMK-MLC}$=8.74) in table 1, this may indicate that the PTJ topic could elicit longer clauses than the SMK topic. This result is consistent with that found in Yoon's (2017) study, where MLC were found to be significantly affected by topics (with an effect size η_p^2=.15 >.1379). Two measures, CP/C and CP/T, of *amount of coordination* are found to be affected. Taking the means of CP/C and CP/T into account (mean$_{PTJ-CP/C}$=0.18 > mean$_{SMK-CP/C}$=0.13; mean$_{PTJ-CP/T}$=0.28 > mean$_{SMK-CP/T}$=0.22), it may indicate the PTJ topic could elicit more coordination expression than the SMK topic. All of the measures, CN/C, CN/T and VP/T in *particular structure* are shown to be significantly different. Comparing their means, it is found that the PTJ topic could elicit more particular structure than the SMK topic.

In ENS group, the only two types of measures are found to be significantly different across topics, which are *length of particular unit* and *particular structure*. Taking the means into account, it can be inferred that the PTJ topic could elicit longer sentences, clauses and T-units and more particular structures than the SMK topic. Compared with CHN group in the type of *amount of coordination*, ENS group is not affected by different topics.

4. Discussion

Comparing the CHN and ENS writings in the same topic, it is found that the ENS writings are generally more complex than the CHN writings in all aspects of length of production units, sentence complexity, amount of subordination, amount of coordination and particular structure. This is reasonable because the writings of CHN groups are selected from the Chinese participants with A2 to B2 proficiency level. This finding echoes with what was reported in Hinkel's (2003) study, in which she found that advanced NNS students tended to employ excessively simple syntactic construction. This phenomenon is in line with the acquisition process of second language. For foreign language learners, simple constructions are acquired earlier than complex constructions. In their language production, simple constructions could be extracted more easily than complex constructions. Therefore, when dealing with the same task, especially in a timed writing task, the EFL writers tend to produce simpler construction than the ENS writers.

It is revealed that the two topics have different effects on CHN writings and the ENS writings. For CHN writings, the PTJ topic triggers more coordination and particular construction than the SMK topic; while for ENS writings, the PTK topic elicits longer production unit than the SMK topic. This finding is in line with the previous study (Yoon 2017), which pointed out that the PTJ topic is more relevant to college student's life than the SMK topic. This could explain the fact that most of complexity measures are significantly higher in the PTJ topic than the SMK topic. For most college students, whether to take a part-time job may be an issue to consider in their daily college life. However, smoking banning in a public place which is more related to social issue but not necessarily to daily life may not be seriously considered by every student. This could result in different degree of familiarity with the topic. In the task complexity point of view, the PTJ topic may be more familiar to college students thus could elicit more complex construction in their writings.

5. Conclusion and Implications

This study revealed significant topic effect on measures of *length of production unit* and *particular structure* in both CHN group and ENS group. For CHN group, the *amount of coordination* could vary across different topics. It is also found that the PTJ topic with more familiarity to college students could elicit more complex construction. This result is surprisingly consistent with the previous studies.

The **limitation** of this study is that it fails to examine the interaction effect of topic and language background due to the following reason: the fact that most measures did not pass the normality testing may probably be caused by the small size or random sampling. Therefore, in the future study, it would be better to take into account larger sample and strict random sampling. Besides, it seems that both topics used in this study is common to students, with minor difference in degree of familiarity. To further investigate the familiarity of topic effect, a pretest on participants' familiarity with the topics is suggested.

Despite the limitation mentioned above, this study still gives some **implication** for language teaching. For language learners, familiar topic could trigger more complex constructions and longer language production (writings). Therefore, the language teachers could utilize writing tasks familiar to students to promote their production training, or provide more information for students to be familiar with the topic before writing.

References

Ai, H., & Lu, X. F. (2013). *A corpus-based comparison of syntactic complexity in NNS and NS university students' writing*. Automatic Treatment and Analysis of Learner Corpus Data. 249-264.

https://doi.org/10.1075/scl.59.15ai

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: which measures? which genre?. *Reading & Writing, 22*(2), 185-200. https://doi.org/10.1007/s11145-007-9107-5

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics, 37*(5), 639-668. https://doi.org/10.1093/applin/amu059

Chen, H. (2010). Study on measuring types and features of English writing performance. *Modern Foreign Languages, 33*(1), 72-80. https://doi.org/10.5771/0170-5067-2010-5-72

Crowhurst, M. (1980). Syntactic complexity and teachers' quality ratings of narrations and arguments. *Research in the Teaching of English, 14*(3), 223-231.

Hinkel, E. (2003). Simplicity without elegance: features of sentences in L1 and L2 academic texts. *TESOL Quarterly, 37*(2), 275-301. https://doi.org/10.2307/3588505

Ishikawa, S. (2013). The ICNALE and Sophisticated Contrastive Interlanguage Analysis of Asian Learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1* (pp. 91-118). Kobe, Japan: Kobe University.

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. https://doi.org/10.1093/applin/amp044

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing, 28*, 53-67. https://doi.org/10.1016/j.jslw.2015.02.002

Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: issues of topic, proficiency, and construct multidimensionality. *System, 66*, 130-141. https://doi.org/10.1016/j.system.2017.03.007

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly, 51*(2), 275-301. https://doi.org/10.1002/tesq.296

Zhao, J. H., & Chen, H. Y. (2012). Towards the measurement of the grammatical complexity in the writings of the learners of English. *Foreign Language Learning, Theory and Practice 1*(1), 27-33.

Notes

Note 1. The web-based L2 Syntactic Complexity Analyzer is available at http://aihaiyang.com/software/l2sca/.

Copyrights