



10.2478/topling-2019-0005

Exploring syntactic complexity and its relationship with writing quality in EFL argumentative essays

Farzaneh Shadloo

Ferdowsi University of Mashhad, Iran

Hesamoddin Shahriari Ahmadi*

Ferdowsi University of Mashhad, Iran

Behzad Ghonsooly

Ferdowsi University of Mashhad, Iran

Abstract

To predict syntactic complexity in second/foreign language writing, some studies have advocated the use of T-unit and clausal subordination measures while others have argued for the use of phrase-based measures. This study seeks to identify syntactic features that can be regarded as discriminators among different levels of writing quality. For this purpose, a corpus of argumentative essays by EFL learners was compiled and then the essays were rated and placed into three groups of high-rated, mid-rated, and low-rated essays. The corpus was then coded and analysed for both phrasal and clausal features. The phrasal features were manually coded based on the development scheme hypothesized by Biber, Gray and Poonpon (2011) for academic writing, and the clausal features were analysed using the online L2 Syntactic Complexity Analyzer developed by Lu (2010). A separate ANOVA test was used to compare the three groups of essays for each of the phrasal and clausal features. The findings of the current study demonstrated that subordination and dependent clauses were not good indicators of different writing qualities in our corpus. Also, the pattern of noun phrase complexity predicted by Biber et al. (2011) was not observed across argumentative essays from three different levels of writing quality.

Key words

corpus linguistics, phrasal features, argumentative essays, EFL learners, T-unit

1. Introduction

In recent years, a great amount of attention has been paid to analysing syntactic complexity in academic writing. Many of these studies have considered T-units and clausal subordination as the most optimal measures for assessing syntactic complexity in academic registers (Biber et al., 2011). Wolfe-Quintero, Inagaki and Kim (1998) refer to dependent clauses per clause and clauses per T-unit as the best criteria for measuring syntactic complexity in written discourse. Similarly, Beers and Nagy (2010) have considered clauses per T-unit and words per clause as two measures of syntactic complexity for exploring the

* Address for correspondence: Hesamoddin Shahriari Ahmadi, Department of English and Applied Linguistics, Ferdowsi University of Mashhad, Azadi Square, Mashhad, Iran. E-mail: h.shahriari@um.ac.ir

differences across four genres of texts including narrative, descriptive, compare/contrast and persuasive writing by a number of students from three different grades.

However, such clause-based views of assessment were challenged in Biber et al.'s study (2011). The authors of this study advocate the use of "complex noun phrase constituents" and "complex noun phrases" as the main criteria for assessing syntactic complexity in L2 writing development (Biber et al., 2011). They argue that as students reach higher levels of proficiency, their writing demonstrates greater levels of complexity, but in a way that cannot be measured via T-unit-based analysis and simple subordination. This, they argue, explains why previous developmental studies have failed to consistently identify a steady rise in the clause-based measures across increasing levels of proficiency (Biber et al., 2011). In support of this claim, Taguchi, Crawford and Wetzel (2013) compared two types of essays (i.e. high- and low-rated essays) written by a number of international students in a university in the United States. By analysing both clause-level and phrase-level complexity measures, the authors supported three claims that (1) highly proficient writers do not necessarily make use of more clausal-level complexity in their essays, (2) subordination cannot be seen as the only measure of complexity in students' writings, and (3) dependent clauses as a measure of complexity cannot be said to represent syntactic complexity in academic writing (Taguchi et al., 2013). Lu (2011) has also referred to the "stronger discriminative power of the complex nominal and coordinate phrase measures than most other measures" (p. 57). He thus calls for more attention in future studies to the examination of complexity at the phrasal level (Lu, 2011).

Argumentative essays are an important register of second/foreign language writing. Such essays are included in the writing component of many standardized English proficiency tests and are featured extensively as assignments in many university programmes. The importance of argumentative essays could be ascribed to the assumption that they can serve as evidence for a student's ability to write in an academic style. Essays have been considered in previous studies as an early-stage representation of academic writing (see Friginal, Li and Weigle, 2014; Kyle, 2016; Parkinson and Musgrave, 2014), and are included in many language proficiency tests, such as the Academic IELTS and the TOEFL iBT, which are required for college admission. However, the question of how well argumentative prose could be said to represent academic writing competence remains to be answered. Investigations into syntactic complexity in argumentative writing could possibly further our understanding of how patterns of complexity vary across writing quality in early stages of academic writing. Specifically, research is needed to show the degree to which argumentative writing follows the same pattern of growth in syntactic complexity as academic registers such as the research article (Biber et al., 2011; Lu, 2011; Parkinson and Musgrave, 2014; Taguchi et al., 2013). Therefore, this study seeks to address this problem by analysing patterns of clausal and phrasal features across argumentative writing from three different levels of writing quality.

To explore patterns of syntactic complexity, the features related to this concept were observed in rated essays from different levels of quality (i.e., low-rated, mid-rated, & high-rated). Attempts were also made to control two variables, namely genre and timing condition, which Lu (2011) found to influence the relationship between proficiency and syntactic complexity. These variables were not controlled in many previous studies into L2 writing development (Lu, 2011).

The present study is guided by the following research questions:

1. Is there a difference in the frequency of those noun modifiers proposed by Biber et al. (2011) in EFL argumentative essays of low-, mid- and high-rated essays?
2. Is there a difference in the frequency of those syntactic complexity measures proposed by Lu (2011) in EFL argumentative essays of low-, mid- and high-rated essays?

2. Review of literature

2.1 Syntactic complexity in writing

Tracking the linguistic changes in L2 writers for the purpose of understanding how L2 writing quality develops with experience is of great importance, and for this reason different measures of complexity in L2 writing such as lexical density, text length, lexical variation and syntactic complexity have to this day been the focus of different studies (Kim, 2014). Among these features, syntactic complexity has traditionally been considered the main index for analysing complexity in L2 writing development (Kim, 2014). Similarly,

Ortega (2003) has argued that “measures of syntactic complexity are important research tools not only in the field of second language acquisition but in a variety of language-related disciplines” (p. 492).

Different linguistic features have been considered as the best measures for measuring syntactic complexity in writing. Most studies have depended on “quantitative variables that measure the average length of structural units or the extent of clausal subordination, assuming that longer units and more subordination reflect greater complexity” (Biber et al., 2011, p. 7). In particular, T-unit-based measures have been widely used in many previous studies of L1 and/or L2 writing. A T-unit is defined as “one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it” (Hunt, 1970, p. 4).

Numerous attempts have been made to find the best measures for identifying syntactic differences among L2 learners of different writing ability levels. Grant and Ginther (2000), for example, analysed 90 essays written by three groups of writers at different proficiency levels. Their findings showed that with an increase in proficiency level, “L2 writers incorporate more subordination, especially complementation, into their timed essays” (Grant and Ginther, 2000, p. 140). In addition to this, the authors found that more nominalizations were used in more advanced essays (Grant and Ginther, 2000). Complex nominalizations have also been pointed out as good indicators of syntactic complexity in writing (e.g., Kim, 2014). In another study of this kind, Ai and Lu (2013) analysed four dimensions of syntactic complexity (i.e., the use of subordination, coordination, phrasal sophistication and length of production unit which includes mean length of T-unit, mean length of clause, and mean length of sentence) in 600 essays written by three groups of writers including two groups of non-native speakers (NNS) with low and high proficiency levels and one group of native speakers (NS). The authors reported significant differences between native- and non-native speakers in terms of their use of dependent clauses per T-unit and dependent clauses per clause, two measures reflecting the amount of subordination. The authors also stated that despite a rise in the use of subordination from non-native-low group to non-native-high group, the difference observed was not statistically significant (Ai and Lu, 2013). As for phrasal sophistication (i.e., complex nominals per clause and complex nominals per T-unit), they observed a significant rise in two features related to phrasal elaboration between the NNS-low group and the NS group and also between the NS group and the NNS-high group (Ai and Lu, 2013). They also found an increase in the measure of complex nominals per T-unit from the non-native-low group to the non-native-high group (Ai and Lu, 2013).

In their seminal study into syntactic complexity in spoken and written English, Biber et al. (2011) argued that T-units and clausal subordination cannot effectively discriminate different levels of language proficiency. The authors of this study maintained that the common measures of syntactic complexity in writing development are in fact characteristic of conversation, and so they introduce features which they believe to be more appropriate for measuring complexity in writing development (Biber et al., 2011). Biber et al. (2011) compared the frequency of 28 grammatical features in conversation with their frequency in academic research articles and found that clausal subordination is a much more characteristic feature of conversation than academic writing. Meanwhile, they discovered that phrasal embedding is the most effective measure of syntactic complexity in academic writing (Biber et al., 2011). On the other hand, “finite clauses functioning as adverbials (if and because) and finite clauses functioning as verb complements (that, ZERO, and WH)”, were more frequently found in conversation (Biber et al., 2011, p. 24). Biber et al. (2011) also pointed out two serious problems with the T-unit, which include “its lack of utility in testing applications [and] its poor theoretical linguistic basis” (p. 12), and concluded by proposing a hypothesized developmental index (see Table 1) which represents how different features of complexity develop as the proficiency level of learners increases.

Table 1. Modified hypothesized developmental stages of noun phrase complexity proposed by Biber et al. (2011)

Stages	Grammatical Structures	Examples
2	^a Simple phrasal embedding in the noun phrase: attributive adjectives	<i>controversial issues</i>
3	^a <i>That</i> relative clauses especially with animate head nouns	<i>citizens that would no longer break any laws</i>
	Simple phrasal embedding in the noun phrase:	
	^b Nouns as premodifiers	<i>imprisonment rates</i>
	^c Possessive nouns as premodifiers	<i>taxpayers' money</i>
	^d Of phrases as postmodifiers	<i>followers of rehabilitation</i>
	^e Simple PPs as postmodifiers, especially with prepositions other than 'of' when they have concrete/locative meanings	<i>chance for rehabilitation</i>
4	^a Nonfinite relative clauses	<i>some problems observed in the system of prison</i>
	^b More phrasal embedding in the NP = attributive adjectives, nouns as premodifiers	<i>high education degree</i>
	^c Simple PPs as postmodifiers, especially with prepositions other than 'of' when they have abstract meanings	<i>attitude towards rehabilitation</i>
5	^a Preposition + nonfinite complement clause	<i>the habit of punishing</i>
	^b Complement clauses controlled by nouns	<i>the fact that colleges and universities try to prepare students</i>
		<i>the opportunity to compensate for his wrongdoings</i>
	^c Appositive noun phrases	<i>So students memorize these material (definitions and formulas) because they did not experience them.</i>
	^d Extensive phrasal embedding in the NP:	
	Multiple prepositional phrases as postmodifiers, with levels of embedding	<i>a result of their separation from the family</i>

In parallel with Biber et al.'s (2011) study, Lu (2011) classified 14 syntactic complexity measures under five headings and analysed them in a corpus of English essays written by college-level Chinese students. These two frameworks had not been analysed altogether in previous L2 writing studies. Table 2 lists the features in question along with their definitions.

Table 2. Syntactic complexity measures adapted from Lu's (2011) study

Measures	Definition
Type 1: Length of production	
Mean length of clause	Number of words divided by number of clauses
Mean length of sentence	Number of words divided by number of sentences
Mean length of T-unit	Number of words divided by number of T-units
Type 2: Sentence complexity	
Clauses per sentence	Number of clauses divided by number of sentences
Type 3: Subordination	
Clauses per T-unit	Number of clauses divided by number of T-unit
Complex T-units per T-unit	Number of complex T-units divided by number of T-units
Dependent clauses per clause	Number of dependent clauses divided by number of clauses
Dependent clauses per T-unit	Number of dependent clauses divided by number of T-units
Type 4: Coordination	
Coordinate phrases per clause	Number of coordinate phrases divided by number of clauses
Coordinate phrases per T-unit	Number of coordinate phrases divided by number of T-units
T-units per sentence	Number of T-units divided by number of sentences
Type 5: Particular structures	
Complex nominals per clause	Number of complex nominals divided by number of clauses
Complex nominals per T-unit	Number of complex nominals divided by number of T-units
Verb phrases per T-unit	Number of verb phrases divided by number of T-units

Lu (2014) concluded that the best measures of syntactic complexity for distinguishing different levels of proficiency “are complex nominals per clause (CN/C) and mean length of clause (MLC), both of which not only discriminated two or more adjacent levels but also increased linearly across all four levels” (p. 56). However, the number of dependent clauses per clause, clauses per sentence, and dependent clauses per T-unit dropped significantly when the learners’ proficiency level increased (Lu, 2014). Consequently, Lu (2014) noted that as students become more proficient, phrasal rather than clausal complexity is more prominently featured in their essays. Despite their differences in measuring the features of complexity, Lu (2014) and Biber et al.’s (2011) studies seem to converge on some major points, most notably that phrasal complexity is a better measure of syntactic development in learner writing. In spite of this, Biber et al. (2011) call for more research into their proposed developmental index.

Parkinson and Musgrave (2014) tested Biber et al.'s (2011) developmental index of different noun modifications by examining the relationship between noun modifications and L2 writing proficiency and confirmed their hypothesis. They analysed Biber et al.'s (2011) phrasal complexity features across two groups of L2 learners including MA students and EAP students and the results were compared against the frequency of noun modifiers in published academic prose. The writing of the EAP group included argumentative essays, whereas the "MA students were writing in response to a variety of extended questions in the field of TESOL/Applied Linguistics" (Parkinson and Musgrave, 2014, p.52). In line with Biber et al.'s (2011) developmental stages of noun modifiers, Parkinson and Musgrave (2014) found that the EAP group, which was assumed to be less proficient than the MA group, mostly used attributive adjectives which are in the lowest stage of Biber et al.'s developmental index of noun modifiers. In contrast, the rest of phrasal noun modifiers incorporating nouns, participial adjectives, prepositional phrases, appositive noun phrases, and possessive nouns were used mostly by the MA group who were believed to be more proficient in academic writing (Parkinson and Musgrave, 2014).

Taguchi et al. (2013) also examined two types of placement exam essays (i.e. high-rated and low-rated essays) by international students from eight different L1 backgrounds. Students' performance was assessed based on language use and content. The former involved the examination of complexity at both clausal and phrasal levels and finally after considering all clausal complexity features, the authors concluded that the higher rated essays were less "complex" than the lower rated essays (Taguchi et al., 2013). On the other hand, among all the features considered as indicators of phrasal complexity (pre-qualifiers, pre-quantifiers, post-determiners, demonstrative determiners, singular definite and indefinite articles, singular or plural determiners, double conjunctions, attributive adjectives, post-noun-modifying prepositional phrases), attributive adjectives and post-noun-modifying prepositional phrases exhibited the most marked difference between the two groups of essays (Taguchi et al., 2013). Therefore, in line with Biber et al.'s (2011) study, they observed that noun phrase modification features such as attributive adjectives and post-noun-modifying prepositional phrases are more likely to affect essay quality (Taguchi et al., 2013).

A recent longitudinal study by Crossley and McNamara (2014) involved determining the quality of 57 L2 learner essays through time and during the course of studying English. The analysis was conducted with respect to the role of syntactic complexity in learners' essays (Crossley and McNamara, 2014). Once again, in line with Biber et al.'s (2011) findings, the authors mentioned that with increasing proficiency, "L2 writers in this study produced fewer incidences of all clauses, longer noun phrases, less syntactic similarity between sentences, fewer verb phrases, more words before the main verb, and more negation" (Crossley & McNamara, 2014, p. 73).

Following Biber et al.'s (2011) developmental index of complexity, Lan (2014) also examined 14 noun modifiers in two sets of essays written by 18 international students. The results of Lan's (2014) study revealed that essays written by the more proficient group of learners included more noun modifiers, and that in general, pre-modifiers were more frequent than post-modifiers in the essays of both groups. The author further argued that postmodifiers which belong to advanced stages of the hypothesized developmental index by Biber et al. (2011), were mostly observed in essays written by the group with a higher L2 writing proficiency level (Lan, 2014).

Finally, in a Multi-Dimensional analysis, Biber, Gray, and Staples (2014) noted that "phrasal noun modifiers (e.g., premodifying nouns, prepositional phrases, genitive of-phrases, and attributive adjectives) function very differently from clausal complexity features, and as a result, they are much more strongly associated with informational, written task types" (p. 25); but clausal complexity was considered a predominant feature of spoken tasks (Biber et al., 2014). Similarly, Swierzbinska (2014) examined the frequency rates of different noun phrase modifications across five levels of proficiency, namely, elementary, middle school, high school, GED and college-level students. He observed that noun phrases with both pre- and post-modifiers were more frequent in texts written by college-level students and also more cases of postmodifiers were observed in their texts. The studies cited above have examined syntactic complexity in different registers of writing, among learners of different L1 backgrounds using different operational definitions for writing ability. Therefore, there is still a need for further studies into how writing by learners from different ability levels differ in the frequency of clausal and phrasal features.

Also the current study attempted to investigate syntactic complexity by comparing two different approaches; in this regard, Lu (2017) also pointed out that it would be beneficial if researchers integrate different tools for examining syntactic complexity measures in corpus-based L2 writing studies as it has not been observed in previous studies.

2.2 Argumentative Writing

There are four basic modes of non-creative writing: descriptive, narrative, expository, and argumentative (Schultz, 1991). Schultz (1991) notes that in terms of cognitive skills, the argumentative essay is the most demanding among these modes of writing as the “writer must go through a complex process which includes examining a problem, evaluating evidence, generating and testing hypotheses, and redefining them in accordance with new ideas and evidence” (Schultz, 1991, p. 412). The argumentative essay is defined as a type of writing in which a writer “takes a point of view and supports it with either emotional or logical appeals” (Crowhurst, 1990, p. 349). Nippold, Ward-Lonergan, and Fanning (2005) have proposed that this type of writing is a “demanding task that requires the use of complex language to analyze, discuss, and resolve controversies in a way that is clear, convincing, and considerate of diverse points of view” (Nippold et al., 2005, p. 125).

Argumentative writing is believed to be “important both for academic success and for general life purposes...[and] the literate, educated person is expected to be able to articulate a position on important matters so as to persuade colleagues, fellow citizens, governments, and bureaucrats” (Crowhurst, 1990, p. 349). This assumption has led to the inclusion of argumentative essays into many language proficiency and aptitude tests required for gaining admission to universities across the world (e.g., IELTS, TOEFL, GRE, & SAT).

Such findings imply the importance of this type of writing prose in academic settings, and as mentioned by Gardner and Nesi (2012), “many studies of academic writing have taken the research article as a model for the sort of writing that students aspire to produce” (p. 28). Therefore, the need for studies which examine syntactic complexity in students’ writing especially in argumentative essays is more felt.

Concerning the use of syntactic complexity in argumentative essays, a recent study by Lu and Ai (2015) involved examining the effect of different L1 backgrounds on syntactic complexity and they found significant differences in dimensions of syntactic complexity across the EFL learners’ essays. For instance, the results of their study showed that the “Tswana group shows a higher level of clausal complexity than phrasal complexity, while the Chinese group shows a higher level of phrasal complexity than clausal complexity” (Lu and Ai, 2015, p. 26).

So far, however, there has been little discussion about the use of syntactic complexity across different levels of writing quality in argumentative essays which represent academic writing as observed in some studies (e.g. Mazgutova and Kormos, 2015; Wulff and Römer, 2009). Therefore, the present study seeks to address this issue by examining features outlined by Biber et al. (2011) and Lu (2011) in three sets of essays from different levels of writing quality.

3. Methodology

3.1 Material and data collection

The data used in this study consists of a selection of argumentative essays. The essays were randomly chosen from university students majoring in English as a foreign language, while some of the essays from pre-intermediate level students were collected from EFL learners studying at private English institutes. The participants’ age range was 18-30 years and they had not spent time in English-speaking countries, and also they had spent the same number of years of studying English in school. All EFL learners in the present study spoke Persian as their mother tongue, and their consent was obtained through the signing of an agreement form. The information related to the corpus compiled for this study is presented in Table 3.

Table 3. Details of the present study's corpus

Essay quality	Number of essays	Total number of words
Low-rated essays	40	20,347
Mid-rated essays	34	20,416
High-rated essays	30	20,223

All the collected essays in this study were based on one of the prompts mentioned below. These prompts are used in the collection of essays for major learner corpora such as the International Corpus of Learner English (ICLE).

1. *Most university degrees are theoretical and do not prepare students for the real world. They are therefore of little value. Do you agree or disagree with the statement.*
2. *The prison system is outdated; no civilized society should punish its criminals. It should rehabilitate them. Do you agree or disagree with the statement.*

Accordingly, learners were asked to state their opinion on the provided topics by writing an argumentative essay in which they explained their reasons for the opinions they held.

3.2 Procedure

Prior to the analysis of syntactic features, the essays were rated by three trained raters with linguistic training and teaching experience and the criterion for rating the essays was the IELTS writing scoring rubric which can be used to assign band scores on essay quality based on four categories: task achievement, coherence and cohesion, lexical resource, and lastly grammatical range and accuracy. The rubric we used includes nine band scores and in our process of rating, essays with a score of 5 or 5.5 were considered as Group A, essays with a score of 6 or 6.5 Group B, and finally essays with a score of 7 or 7.5 were placed in the category of Group C. It should be taken into account that essays with grade C do not necessarily represent advanced level writing, but rather they are essays with the highest score among the corpus of essays examined in the current study. The score for each essay was the average score assigned to it by the three raters. The three raters underwent a training and calibration phase, in which they rated sample essays together sharing their reasons. Following this, the three raters independently rated a second sample essay and subsequently compared their ratings trying to reach an agreement on how they had assigned their scores. This process increased the level of agreement among the three raters. Regarding inter-rater reliability, ICC estimates and their 95% confident intervals were calculated using SPSS statistical package version 23 (SPSS Inc, Chicago, IL) based on a mean-rating ($k = 3$), absolute-agreement, 2-way mixed-effects model and the 95% confidence interval of the ICC estimate showed 0.79-0.89, which is regarded as "good" reliability. Any essay not tailored to these three levels was excluded from our study.

The three sub-corpora of essays were compared based on the hypothesized indices of complexity in academic writing by Biber et al. (2011) and also the 14 measures of syntactic complexity examined by Lu (2011). Biber et al.'s (2011) framework consisted of attributive adjectives, *that* relative clauses, nouns as premodifiers, possessive nouns as pre-modifiers, *of* phrases as postmodifiers, prepositions other than *of* (with concrete/ locative meanings), nonfinite relative clauses, noun + adjective sequences, prepositions other than *of* (abstract meanings), preposition + nonfinite complement clauses, noun complement clauses, appositive noun phrases, and multiple prepositional phrases as post-modifiers. Following this, the use of 14 measures of syntactic complexity including mean length of clause, mean length of sentence, mean length of T-unit, clauses per sentence, clauses per T-unit, complex T-units per T-unit, dependent clauses per clause, dependent clauses per T-unit, coordinate phrases per clause, coordinate phrases per T-unit, T-units per sentence, complex nominals per clause, complex nominals per T-unit, and verb phrases per T-unit were analysed.

3.2.1 Coding

All features of phrasal complexity in Biber et al.'s (2011) taxonomy were manually hand-coded and their frequency was calculated. Before starting the coding process, all quotations used in learners' essays were replaced with the symbol <R>; this was because the aim of the study was examining learners' language. To

ensure reliability, we also made use of online taggers, such as the CLAWS tagger developed at Lancaster University to inform the process of categorization. Computational procedures used for (semi)-automatic linguistic processing of passages provides complete linguistic characterization of a passage by description of a broad range of linguistic features (Biber, Conrad, Reppen, Byrd and Helt, 2002). This would allow coders to check their categorization of noun modifiers with the automatic POS tags assigned by CLAWS. Doing so would facilitate the coding process by decreasing the chance of mistakes. The coding of NP features was performed by two raters and the inter-rater reliability was 0.91.

The 14 syntactic complexity measures by Lu (2011) were categorized into five main types including length of production, sentence complexity, subordination, coordination, and other structures found. These measures were examined using the online L2 Syntactic Complexity Analyzer designed by Lu (2010). When corpus files are uploaded to the analyser, it “produces frequency counts of nine linguistic units in the text—word, sentence, clause, dependent clause, T-unit, complex T-unit, coordinate phrase, complex nominal, and verb phrase—and generates 14 indices of syntactic complexity for the text” (Yang, Lu and Weigle, 2015, p. 58).

3.3 Data analysis

Following the previous steps, comparisons were made in the present study to find whether there were any statistically significant differences in the frequency of different noun modifiers and the 14 measures of syntactic complexity used by Lu (2011) across the essays from the three different writing qualities. For analysing noun modifiers, all frequency counts were normalized to 600 words (i.e., the average length of an essay) in order to allow for the use of parametric statistical procedures. Following the coding and normalization processes, all data were entered into SPSS.

A one-way ANOVA was run for every single one of the feature mentioned. For those features where the assumption of homogeneity was not met, indicated by the Levene’s Test of Homogeneity of Variances, a non-parametric Welch test was carried out instead. The Tukey Post-Hoc test was used to confirm where the differences occurred between the groups.

The Bonferroni and Tukey tests are both used to control Type I error (Field, 2009). “Of the two, Bonferroni has more power when the number of comparisons is small, whereas Tukey is more powerful when testing large numbers of means” (Field, 2009, p. 374). In addition, through personal correspondence with Xiaofei Lu, the developer of the Online L2 Syntactic Complexity Analyzer software, we were advised to use the Tukey post-hoc test for our data. Effect size, as represented by Cohen’s *d*, is also reported for those features that showed a statistically significant difference.

One final note should be made regarding the statistical analysis of our findings. In this study, we ran 26 separate ANOVA tests to compare syntactic features across three groups of essays. In such cases, a Bonferroni adjustment to the alpha level is required for significance testing. A comparison between results with the alpha level set at 0.05 and an adjusted alpha level of 0.0019 reveals a number of differences which will be presented in the data analysis below. It is important to note here that differences found at the 0.05 level are mere tendencies in the data, and the truly significant differences are those arrived at using the adjusted 0.0019 level of significance.

4. Results

4.1 Noun phrasal complexity

The results of the one-way ANOVA showed that, of the 13 features of noun phrasal complexity, attributive adjectives [$F(2, 101) = 3.28, p = 0.04, d = 0.73$], *that* relative clauses (with animate head nouns) [$F(2, 64.57) = 2.99, p = 0.05, d = 0.59$] and nonfinite relative clauses (*ed*- and *ing*-clauses) [$F(2, 55.23) = 14.16, p = 0.00, d = 1.35$] were the only features that were used in significantly different rates across the three groups of essays. The results of Tukey Post-Hoc showed that regarding attributive adjectives a significant difference was found between group B and C. Concerning *that* relative clauses, a significant difference was found between group A and C, and finally for nonfinite relative clauses, groups A and B as well as groups A and C revealed significant differences. The results of ANOVA statistics in addition to mean scores of each group

of essays are presented in Table 4 where significant differences at the 0.05 and the adjusted 0.0019 level can be found.

Table 4. Results of noun phrasal complexity

Linguistic Feature	ANOVA Statistics	Mean Scores (per 1000 words)			Cohen's <i>d</i>
		Group A	Group B	Group C	
Attributive adjectives	F= 3.28; P= 0.04 (*)	28.66	27.32	33.16	0.73
<i>That</i> relative clauses	F= 2.99; P= 0.05 (*)	0.63	0.18	0.09	0.59
Nouns as premodifiers	F= 0.38; P= 0.68	4.98	5.68	5.06	
Possessive nouns	F= 0.11; P= 0.86	0.54	0.45	0.42	
<i>Of</i> phrases as postmodifiers	F= 0.99; P= 0.37	8.86	9.38	10.34	
Prepositions other than <i>of</i> (concrete/ locative meanings)	F= 1.46; P= 0.23	1.79	2.31	1.56	
Nonfinite relative clauses	F= 14.16; P= 0.00 *	0.54	1.27	2.14	1.35
More phrasal embedding in the NP	F= 1.05; P= 0.35	1.53	0.91	1.10	
Prepositions other than <i>of</i> (abstract meanings)	F= 2.65; P= 0.07	2.53	2.74	3.83	
Preposition + nonfinite complement clauses	F= 1.28; P= 0.28	1.74	1.43	1.87	
Noun complement clauses	F= 2.55; P= 0.08	3.25	4.08	4.58	
Appositive noun phrases	F= 0.14; P= 0.86	0.21	0.18	0.14	
Extensive phrasal embedding in the NP	F= 0.99; P= 0.37	1.85	1.53	2.09	

Note. (*) shows significance at 0.05, and * shows significance at 0.0019.

Although there was a statistically significant difference in the frequency of attributive adjectives across the three groups, the distribution of this feature did not increase linearly as the quality of essays improved. That is, the only significant difference was between mid-rated and high-rated essays, with the latter using more attributive adjectives than the former.

The frequency of *That* relative clauses, on the other hand, decreased significantly from lower to higher levels. Biber, Johansson, Leech, Conrad and Finegan (1999) have noted that “*that* has more informal, colloquial associations and is thus preferred in conversation” (p. 616). This may be one reason why *that* relative clauses are more frequently found in essays by the less proficient group of this study. Moreover, learners from group C made more use of other forms of relativizers, so fewer cases of *that* relativizers were observed in their essays.

The analysis of nonfinite relative clauses revealed that as the writing quality of essays in this study increased, so did the use of nonfinite relative clauses. This significant difference among the three groups is in line with Biber et al. (1999), who found that nonfinite relative clauses (*ed*-clauses and *ing*-clauses) are more frequently used in academic prose compared to news, fiction and conversation registers.

Considering Biber et al.’s (2011) taxonomy, only three features could be considered as useful discriminators among different levels of writing quality. The other features in Biber et al.’s (2011) taxonomy did not follow patterns that were of statistical significance and they could not be used to predict syntactic differences across essays of different levels of quality.

In line with the findings of this study, Biber et al. (2014) found that the hypothesized stages of noun modifications proposed in Biber et al. (2011) could not be observed when analysing a corpus of standardized TOEFL iBT writing responses. The authors of that study concluded that when considering proficiency differences, “analyses based on individual complexity features generally failed to identify significant linguistic differences across score levels” (p. 26) and they pointed out that “one complication here is that exam scores reflect many factors in addition to grammatical proficiency” (p. 8).

4.2 Lu's (2011) proposed syntactic complexity measures

As for syntactic complexity measures, the results of the one-way ANOVA showed significant differences ($p < 0.05$) in the mean values of five measures, including mean length of T-unit [$F(2, 101) = 5.43, p = 0.00, d = 0.93$], mean length of clause [$F(2, 101) = 5.58, p = 0.00, d = 0.94$], verb phrase per T-unit [$F(2, 101) = 5.61, p = 0.00, d = 0.93$], complex nominals per T-unit [$F(2, 101) = 3.67, p = 0.02, d = 0.77$], and complex nominals per clause [$F(2, 101) = 4.23, p = 0.01, d = 0.81$] across Iranian EFL learners' argumentative essays. Post-Hoc results demonstrated significant differences between high- and low-rated essays for all the features that results of one-way ANOVA indicated statistically significant differences. The results of ANOVA statistics in addition to mean scores of each group of essays are provided in Table 5. Significant differences at the 0.05 and 0.0019 levels can be seen in the table.

Table 5. Results of Lu's (2011) proposed syntactic complexity measures

Linguistic feature	ANOVA statistics	Mean Scores (per 1000 words)			Cohen's <i>d</i>
		Group A	Group B	Group C	
Mean length of sentence (MLS)	$F=2.58; P= 0.08 > 0.05$	20.51	21.47	23.78	
Mean length of T-unit (MLT)	$F=5.43; P= 0.00 *$	16.20	17.73	19.71	0.93
Mean length of clause (MLC)	$F=5.58; P= 0.00 *$	9.42	10.02	11.04	0.94
Clauses per sentence (C/S)	$F=0.24; P= 0.78$	2.17	2.14	2.22	
Clauses per T-unit (C/T)	$F=1.79; P= 0.17$	1.71	1.75	1.84	
Verb phrases per T-unit (VP/T)	$F=5.61; P= 0.00 *$	2.30	2.48	2.72	0.93
Dependent clauses per clause (DP/C)	$F=3.25; P= 0.04$	0.36	0.38	0.41	
Dependent clauses per T-unit (DP/T)	$F=1.96; P= 0.14$	0.65	0.69	0.76	
T-unit per sentence (T/S)	$F=0.95; P= 0.38$	1.26	1.20	1.24	
Complex T-unit ratio	$F=2.82; P= 0.06$	0.46	0.48	0.54	
Coordinate phrases per T-unit (CP/T)	$F=0.15; P= 0.86$	0.51	0.54	0.56	
Coordinate phrases per clause (CP/C)	$F=0.16; P= 0.84$	0.29	0.30	0.31	
Complex nominal per T-unit (CN/T)	$F=3.67; P= 0.02 (*)$	1.94	2.08	2.36	0.77
Complex nominal per clause (CN/C)	$F=4.23; P= 0.01 (*)$	1.12	1.17	1.32	0.81

Note. (*) shows significance at 0.05, and * shows significance at 0.0019

As shown in Table 5, two out of three measures which were related to the length of production and all three measures of CN/T, CN/C, and VP/T increased significantly and linearly as the level of essay quality rose. Except verb phrases per T-unit, the other four measures were similarly found to increase in statistically significant terms in essays investigated in Lu's (2011) study. This observation corroborates the findings of Yang et al. (2015) who noted that the mean length of clause and mean length of T-unit can be used in "examinations of the relationship between syntactic complexity and writing quality" (p. 63).

5. Discussion

The present study analysed syntactic complexity in essays with three different levels of writing quality. The aim of the analysis was to identify measure(s) that could best predict differences across these levels of writing. Phrasal features outlined by Biber et al. (2011) and Lu's (2011) comprehensive set of indices were analysed for this purpose. The findings showed that from Biber et al.'s (2011) taxonomy, three features (i.e.,

attributive adjectives, *that*-relative clauses and non-finite relative clauses) and from Lu's (2011) framework, five features predicted syntactic change across different writing qualities. It should be noted, however, that the only feature for which true statistical significance (at the 0.0019 level) was found was finite relative clauses in Biber's (2011) taxonomy. The other mentioned differences could only be taken as tendencies of variation in the data.

Concerning Lu's (2011) framework, the findings of our study revealed that subordination and dependent clauses cannot be regarded as powerful discriminators of different writing qualities. Moreover, it can be noted from our results that high-rated essays contained longer clauses and T-units and the learners achieved this by making use of more complex nominals. Then regarding features of noun-phrasal complexity, it is shown that attributive adjectives, *that* relative clauses and non-finite clauses could to some extent distinguish different writing qualities in argumentative essays. Thus, by analysing the concept of syntactic complexity via two different measurements, a more comprehensive picture was formed. In line with this explanation, Norris and Ortega (2009) also pointed out that there is a need for other studies of L2 writing to examine syntactic complexity at a more general level by measuring complexity at phrasal level and complexity via subordination.

There can be some possible explanations for why the results of the present study do not follow the same pattern of development proposed by Biber et al. (2011). One possible explanation is that the stages in the hypothesized index of complex noun modifications in Biber et al.'s (2011) study emerged from analysing a large sample of research articles written by professional writers. Based on this data, the authors hypothesized that their index may also be applied to L2 learners of English. However, the findings of the present study as well as those of Biber et al. (2014) suggest that perhaps syntactic development in argumentative essays follows a pattern different from that of academic research articles.

The developmental stages of Biber et al. (2011) could not be used to predict changes that occur at the syntactic level across ability levels in this study, which shows that, at least at a syntactic level, argumentative essays may not accurately represent academic writing ability. Although both the argumentative essay and the research article are considered to be common written registers in the university context, the same syntactic complexity features found in a corpus of research articles may not correspond to those within the register of argumentative essays. This could be attributed to the varying standards of what constitutes "good" writing within the general English classroom and the higher education setting. In many parts of the world, language learners are encouraged to lengthen their sentences and compose what is believed to be "complex" and "elaborated" constructions. This is while clarity of expression is valued in academic settings and university students are prompted to work towards this aim in both their written assignments and academic publications.

This issue raises an important concern for English language teachers and specifically English writing instructors, who wish to make their students familiar with identifying linguistic characteristics of varying registers and even sub-registers.

The findings discussed here also have useful implications for developers of materials for writing instruction to take into account the differences in the patterns of syntactic complexity among sub-registers of academic writing. Also, it would be useful for researchers to look for grammatical structures that characterize other types of writing registers and sub-registers, especially those which learners commonly deal with. This can help to provide a more accurate and realistic assessment of learners' writing in different contexts.

One limitation of this study is that our corpus of argumentative essays are written based on two topics. Adding to the size of the corpus and perhaps including students from other L1 backgrounds could possibly strengthen the generalizability of the results. Also, caution must be applied as our findings cannot be used to make predictions regarding students' proficiency, but rather only the quality with which the essays are written. It should also be noted that many of the differences found between the findings of this study and those of Biber et al. (2011) might possibly be due to register differences between argumentative essays and academic subgenres such as the research article. Further, dividing up the corpus into three levels and not considering other levels of ability beyond this defined window is a limitation that could be overcome by future studies into this area.

Future studies are also recommended to analyse syntactic complexity in the essays of different L1 backgrounds via different types of metrics. Moreover, researchers could conduct studies into other features (e.g., lexical, discoursal, etc.) of the argumentative essays in order to determine whether this register could indeed be used to represent academic writing.

Acknowledgments

We would like to express our gratitude to the reviewers and especially the editor of the journal *Topics in Linguistics* for their careful reading of our manuscript and their valuable comments and suggestions for improving the quality of our study.

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*. Amsterdam/ Philadelphia: John Benjamins, pp. 249–264.
- Beers, S. and Nagy, W., 2010. Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, vol. 24, no. 2, pp. 183–202.
- Biber, D., Conrad, S., Reppen, R., Byrd, P. and Helt, M., 2002. Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, vol.36, no.1, pp. 9–48.
- Biber, D., Gray, B. and Poonpon, K., 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, vol. 45, no. 1, pp. 5–35.
- Biber, D., Gray, B. and Staples, S., 2014. Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E., 1999. *Longman grammar of spoken and written English*. Harlow, England: Longman.
- Crossley, S. A. and McNamara, D. S., 2014. Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, vol. 26, pp. 66–79.
- Crowhurst, M., 1990. Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education / Revue Canadienne De L'éducation*, vol.15, no. 4, pp. 348–359.
- Field, A., 2009. *Discovering statistics using SPSS*. Sage publications.
- Friginal, E., Li, M. and Weigle, S. C., 2014. Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, vol. 23, pp. 1-16.
- Gardner, S. and Nesi, H., 2012. A classification of genre families in university student writing. *Applied Linguistics*, vol. 34, no. 1, pp. 25–52.
- Grant, L. and Ginther, A., 2000. Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, vol. 9, pp. 123–145.
- Hunt, K. W., 1970. Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development*, vol. 35, no.1, pp. iii–67.
- Kim, J., 2014. Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching*, vol. 69, no. 4, pp. 27–51.
- Kyle, K., 2016. Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication (Doctoral dissertation). Retrieved from http://scholarworks.gsu.edu/alesl_diss/35.
- Lan, G., 2014. *Investigating the relationship between second language writing proficiency and noun modification*. Paper presented at Second Language Writing Symposium of Arizona State University, Arizona, United States.
- Lu, X., 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, vol. 15, pp. 474–496.
- Lu, X., 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, vol. 45, no. 1, pp. 36–62.

- Lu, X., 2017. Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, vol. 34, no. 4, pp. 493–511.
- 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*, vol. 29, pp. 16–27.
- Mazgutova, D. and Kormos, J., 2015. Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, vol. 29, pp. 3–15.
- Nippold, M. A., Ward-Loneragan, J. M. and Fanning, J. L., 2005. Persuasive writing in children, adolescents, and adults. *Language Speech and Hearing Services in Schools*, vol. 36, no. 2, pp.125–138.
- Norris, J. M. and Ortega, L., 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, vol.30, no. 4, pp. 555–578.
- Ortega, L., 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, vol. 24, no. 4, pp. 492–518.
- Parkinson, J. and Musgrave, J., 2014. Development of noun phrase complexity in the writing of English for academic purposes students. *Journal of English for Academic Purpose*, vol. 14, pp. 48–59.
- Schultz, J. M., 1991. Writing mode in the articulation of language and literature classes: Theory and practice. *The Modern Language Journal*, vol. 75. no. 4, pp. 411–417.
- Swierzbin, B., 2014. What's in a noun phrase? Judging the difficulty of a reading text by understanding the complexity of noun phrases. *Minnetesol - Minnesota Teachers of English to Speakers of Other Languages*. Retrieved from <http://hdl.handle.net/11299/176674>.
- Taguchi, N., Crawford, W. and Wetzel, D. Z., 2013. What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, vol. 47, no. 2, pp. 420–430.
- Wolfe-Quintero, K., Inagaki, S. and Kim, H. Y., 1998. *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: University of Hawaii Press.
- Wulff, S. and Römer, U., 2009. Becoming a proficient academic writer: Shifting lexical preferences in the use of the progressive. *Corpora*, vol. 4, no. 2, pp.115–133.
- Yang, W., Lu, X. and Weigle, S. C., 2015. Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgements of writing quality. *Journal of Second Language Writing*, vol. 28, pp. 53–67.