

# Lexical and Syntactic Complexities in Undergraduate Students' Research Articles and Their Correlations to Their Quality

**Ratna Dewi**

Universitas Muhammadiyah Makassar, South Sulawesi, Indonesia  
Jl. Sultan Alauddin No. 256, Makassar 90223, Indonesia. E-mail: [iratnadewi@yahoo.com](mailto:iratnadewi@yahoo.com)

**Abstract:** This study aims to find out the employment of lexical and syntactic complexities in the undergraduate students' research articles and their correlations to their quality. It employs a quantitative design through corpus based analysis and correlation analysis using Pearson product-moment correlation coefficient. The results show that lexical density, lexical sophistication, and lexical variation are high except for verb variation, noun variation, adjective variation, adverb variation and modifier variation and the sentence complexity is also high. The quality of the undergraduate students' research articles is categorized as above good or it is only a half point needed to achieve the criteria of great research article. This research find that there is no correlation between lexical and syntactic complexities with the quality of articles. It is expected that this article could provide information about the lexical and syntactic complexities which are needed to be improved in the undergraduate students' research articles.

**Key Words:** lexical complexity, syntactic complexity, article quality

**Abstrak:** Penelitian ini bertujuan untuk mengetahui penggunaan kompleksitas leksikal dan sintaksis pada artikel penelitian mahasiswa dan korelasinya terhadap kualitasnya. Penelitian ini menggunakan desain kuantitatif melalui analisis berbasis korpus dan analisis korelasi menggunakan koefisien korelasi momen produk Pearson. Hasil penelitian ini menunjukkan bahwa kepadatan leksikal, kecanggihan leksikal, dan variasi leksikal tinggi kecuali pada variasi verba, variasi kata benda, variasi kata sifat, variasi adverbial dan variasi pengubah dan kompleksitas kalimat juga tinggi. Kualitas artikel penelitian mahasiswa tingkat sarjana dikategorikan sangat baik atau hanya setengah poin yang dibutuhkan untuk mencapai kriteria penelitian besar. Penelitian ini menemukan bahwa tidak ada korelasi antara kompleksitas leksikal dan sintaksis dengan kualitas artikel. Diharapkan artikel ini dapat memberikan informasi tentang leksikal dan kompleksitas sintaksis yang perlu ditingkatkan dalam artikel penelitian mahasiswa.

**Kata kunci:** kompleksitas leksikal, kompleksitas sintaksis, kualitas artikel

Lexical and syntactic complexities characterize academic written texts of advanced writers. The high proficient writers emerge with the more sophisticated vocabularies, those are Base Word 3, the University Word List (UWL) and 'not in any list words (Laufer & Nation, 1995, p. 316). Students of higher proficiency level tend to produce longer and complex sentences and longer clauses and T-units in the forms of complex phrases such as coordinate phrase and complex nominals (Lu, 2010; Yau & Belanger, 1984; Mukminatun, 1997).

Lexical complexity use describes the writer's ability to communicate effectively in written form and

syntactic complexity use delineates the writer's overall sentence development in the target language (Lu, 2010; Lu, 2012; Ai & Lu, 2010). Due to the fact, the existence of lexical and syntactic complexities in students' academic texts sets forth the students' writing proficiency. Therefore, lexical and syntactic complexity proficiency in writing academic texts such as research articles is undoubtedly required.

The appearance of lexical and syntactic complexities in academic text is also the nature of the text itself that loads complex ideas, which need lexical and syntactic complexities to generate them meaningfully. The complex ideas can be more flexibly and meaning-

fully explained through the wide range of vocabulary use, and can be specifically and sophisticatedly generated through the use of specific words, which are found in Base Word 3, in 'not in any list and University Word List. Moreover, complex ideas are commonly written in complex lexis and sentences in order to accommodate the needs for describing and explaining specification. Pertaining to the nature of academic text, writers, in general, needs to implement lexical and syntactic complexities in their academic texts.

In short, academic texts are characterized by the extensive use of lexical and syntactic complexities. Academic texts including journals or research articles utilize a wide variety of vocabularies, exhibit the use of unusual or advanced words, and label a wide range of vocabulary. Academic texts also rely on longer sentences, syntactic modifiers or subordinate clauses, and complex nominal.

So far, the studies done are mainly focused on the differences of the existence of lexical and syntactic complexities in the students' academic texts of different levels as a result of length of time in learning. The amount of variety and sophistication of the students' lexical and syntactic complexities use increase along with the length of learning and experience in writing (Laufer & Nation: 1995; Lu: 2010, 2012). The students of different proficiency levels in writing are significantly different in their lexical richness (Laufer & Nation, 1995, p. 316). The less proficient students made more use of the first 1,000 most frequent words in their texts. In the other side, high proficient students emerge intensively with the more sophisticated vocabulary, they are Base Word 3, 'Not-in-any-lists' words, and the UWL.

Other research related to Test of Written English explains that lexical and syntactic complexities are one of the important constructs because it can gauge the L2 writers' writing scores (Fraser et al., 1999 in Hinkel, 2003; Francis et al., 2002). The scores are given based on the extent of word type used in the text, the intensive use of advanced or derived words (unique and longer words) and the proportion of content words exhibited in the text. Moreover, Hinkel (2003, p. 276) states that the degree of sophistication of text determined by syntactic complexity is identified through the extensive use of subordinate clauses. The words and sentences employed by the writers in their writing describe their lexical and syntactic complexities which are the part of language criteria that reflected the writer's proficiency.

Different research related to lexical and syntactic complexities were conducted by Larsen-Freeman (2006) and Naves (2007) who found that Learners who became older, more instructed, and more sophisticated, started neglecting accuracy and fluency and start to concentrate on lexical and syntactic variety. At that time, the learners became more challenged to perform their capacity to use more advanced language. They involved a greater willingness to take risks and to use fewer controlled language subsystems. They were more likely to use more adjective clauses, more modifiers, more complex nominal, as well as gerunds and infinitives.

One different research of different time on lexical and syntactic complexities was done by Hinkel (2003, 2005, and 2011) and Sylva (1993) who described lexical and syntactic complexities of L2 writers' academic texts by comparing them with the native writer's text. Hinkel (2003, p. 297) stated that NNSs' productive range of grammar and lexis was comparatively small and consisted largely of construction, prevalent in spoken discourse as well as high-frequency, and every day vocabulary items. Hinkel (2005, p. 622) reported that after years the L2 writers continued to differ from that of the novice NS in regard to a broad range of features. She established, however, that even advanced and trained L2 writers had severely limited lexical and syntactic repertoires that enabled them to produce simple texts, restricted to the most common language features in conversational discourse.

In the Indonesia context, the only research done to the written texts was to study the lexical richness or in this study said as lexical sophistication. Afini and Cahyono (2012) found that both male and female students used the 2,000 most frequent words repetitively. In other words, the students' lexical sophistication was considered low since 79.12% of the word families used were included as high frequency words. Meanwhile, the only research related to students' production of syntactic complexity in academic texts was conducted by Mukminatun (1997). The research found that the sentences produced by the students increase through consecutive years.

Considering the vacuity of research on lexical and syntactic complexities, especially the employment of lexical and syntactic complexitie in academic texts, this present research aims to analyze the lexical and syntactic complexities employed in the undergraduate students' research articles and their correlation to the research article quality as evidence for their acquisition

after learning in English Department of Faculty of Letters, Malang State University. This problem is specifically answered by finding the values got by the students in the employment of lexical complexity covering lexical density, lexical sophistication, lexical variation, as well as in the employment of syntactic complexity comprising length of production unit, sentence complexity, amount of subordination, amount of coordination, and degree of phrasal sophistication; by assessing the value of the quality of articles and by finding the correlation of each component of lexical and syntactic complexities to the quality of articles..

### METHOD

Employing a quantitative design through corpus based analysis, this study examined the employment of lexical and syntactic complexities in the undergraduate students' research articles using Lexical Complexity Analyzer (LCA) and Syntactic Complexity Analyzer (L2SCA) and to examine the correlation between the count results of each indicator that signify lexical and syntactic complexities and the values of the undergraduate students' research articles using Pearson Product Moment Correlation.

Lexical Complexity Analyzer (LCA) required file .txt. in which the content had been in the forms of only paragraphs without pictures, graphs, tables, figures, references, title and subtitles. These data had also been ascertained following the American Spelling through the process of scanning. Before the analysis done in LCA, the .txt.files were firstly tagged in Stanford POS Tagger then lemmatized using MORPHA. Next, the output was taken by LCA as input. The output of LCA was in the form of numbers which described the values of each criteria of lexical complexity, namely, lexical density, lexical sophistication, and lexical variation. Lexical density was analyzed using lexical density measure (LD), lexical sophistication using (LS1, LS2, VS1, VS2, CVS1), lexical variation using (NDW, NDW-50, NDW-ER, NDW-ES, TTR, MSTTR, CTTR, RTTR, AdjV, AdvV, ModV).

The same data or similar .txt-files were analyzed to find out the values of syntactic complexity of the research articles. In this part, the .txt.files were parsed in STANFORD PARSER, the outputs of this parser were queried in TREGEX, then finally counted in L2SCA. The output of this analyzer was the count results of length of production unit measures (MLS, MLC, MLT), sentence complexity measure (C/S), amount of subordination measures (C/T, CT/T, DC/

C, DC/T), amount of coordination measures (CP/C, CP/T, T/S), and degree of phrasal sophistication measures (CN/C, CN/T, VP/T).

Since all the outputs from each measurers were in the form of numbers, in order to know the level of complexity of lexis, the values were compared to the values of lexical complexity of Chinese learners' spoken narratives. The same intention was also imparted to the values of the syntactic complexity in the undergraduate students' research articles which were compared to the values of syntactic complexity of argumentative essays of NNS-High of Chinese learners in WECCCL and the values of argumentative essays of NS in LOCNESS.

Next, all the count results of lexical and syntactic complexity measures were correlated with the values of the quality of research articles which were taken from the process assessment of two raters.

A scoring rubric was developed for the purpose of providing similar criteria in scoring the quality of the undergraduate students' research articles. The scoring rubric consisted of four components, they were flawless language, convincing rhetoric, retrieving academic insight, and elegant style. *Flawless language* referred to the absence of grammatical, spelling, and punctuation errors in the research articles. *Convincing rhetoric* pertained to the ability of the writer to provide an argument by showing the importance of the topic, exposing how far the topic had been investigated, stating that the topic was extended and continued from previous research, explicitly mentioning the coverage of the topic, developing a clear and appropriate design, connecting the result with the previous research through comparison, exemplification, explanation, deduction, and recommendation. *Retrieving academic insight* related to the fulfillment of important aspects written in each component of writing, namely, abstract, background, method, result, discussion, conclusion and suggestions, as well as references. *Elegant style* pertained to the extensive and intensive use of lexical and syntactic complexities which preserved to a comprehensible meaning. This scoring rubric had been validated by the expert.

### RESULTS AND DISCUSSION

#### The Undergraduate Students' Lexical Complexity

Lexical Complexity Analyzer (LCA) is design by Prof. Xiaofei Lu from Pennsylvania University, USA

for measuring lexical complexity of spoken and written texts. The oral narratives produced by the test takers are transcribed into files kept in SECCL (Wen, Wang, & Liang, 2005 in Lu, 2010). The LCA is assigned to measure the written text files originated from oral texts as well as from written texts. Notably, the LCA is administered to identify the lexis accepted as input without noticing the original one, as long as the input has been in the form of acceptable written text files for the system. The LCA just focuses on identifying the lexis accepted as input and revealed the count results of the lexis as output.

It should be noted that both spoken and written texts consist of lexis as small components of texts that contains meaning. Since the contents are similar, namely, words, the researcher considers the comparison to be viable for the sake of defining how high the complexity of the students' lexis. In this context, spoken and written data are similar (Ure, 1971 & O'Loughlin, 1995 in Lu, 2012; Brown, 2007). Comparing spoken and written texts has been done by some researchers who reports that spoken texts have a lower lexical density than written texts (Ure, 1971 & O'Loughlin, 1995 in Lu, 2012; Brown, 2007). If the comparison in those researches is referred to this report, the consequences for the result of comparison which have similar count results between spoken and written texts should be directed to the more improvement of lexical complexity in the students' research articles. Based on the explanation, the researcher compares the count results of the lexical complexity of undergraduate students' research articles and the lexical complexity of Chinese learners' oral narratives.

**Lexical Density in Undergraduate Students' Research Articles**

Lexical complexity in the students' research articles is characterized by the presence of three features: lexical density, lexical sophistication, and lexical variation (Ai & Lu 2010; Lu, 2012; Siskova, 2012).

The existence of lexical density in the undergraduate students' research articles is shown by the ratio of lexical words compared with the total number of words in the research articles. Lexical words cover nouns, adjectives, verbs, and adverb (Lu, 2012). Table

**Table 1. The Undergraduate Students' Lexical Density**

	Lexical Density Values
Mean Value	0.52
Maximum Value	0.57
Minimum Value	0.48

1 describes the density of the lexical words identified in the undergraduate students' research articles. It informs that the density mean of the lexical words used by the undergraduate students is 0.52 of the total number of words used in the research articles. The representation of the employment of lexical density in the undergraduate students' research articles shown by the mean value obtained are 0.520. This value is higher than that of Chinese oral narratives (0.414). In other words, the number of lexical words implemented in the undergraduate students' research articles are bigger than those in Chinese oral narratives.

**Lexical Sophistication in Undergraduate Students' Research Articles**

Another feature which exhibits lexical complexity of the research articles is the existence of lexical sophistication. Lexical sophistication is advanced words or relatively unusual words in the students' research articles. Five measures accommodated in LCA were used to count lexical sophistication, they were LS1 (Linnarud, 1986; Hylstenstam, 1988); LS2 (Laufer, 1994); VS1 (Harley & King, 1989); VS2 (Choudron & Parker, 1990); CVS1 (Wolfe-Quintero et.al, 1998). LS1 & LS2 measures counted the ratio of the advanced or unusual words to the total number of lexical words in the research articles. VS1 computed the ratio of the advance verbs or the relatively unusual verbs to the total number of lexical verbs in the research articles. VS2 & CVS1 are the same kind of measures that count verb sophistication in the writers' texts but with different formula which are made to reduce the sample size effect of the count. Table 2 shows the values of the lexical sophistication of the undergraduate students' research articles using the five measures.

Description of the employment of lexical sophistication in the undergraduate students' research articles which is described by earlier sophistication measures

**Table 2. The Undergraduate Students' Lexical Sophistication**

	Lexical Sophistication Values				
	LS1	LS2	VS1	VS2	CVS1
Mean Values	0.29	0.32	0.07	2.97	1.09
Maximum Values	0.47	0.48	0.18	20.53	3.20
Minimum Values	0.18	0.22	0.02	0.21	0.32

such as LS1, LS2, and VS1 (0.29, 0.32, & 0.07) are higher or similar compared with the mean values of Chinese Learners' spoken narratives (0.227, 0.262, & 0.074). Meanwhile through the mean value of lexical sophistication using transformed measures such as VS2 (2.97) and CVS1 (1.09), the number of advanced words used by the undergraduate students are bigger than those of Chinese learners' spoken narratives (0.312 & 0.330). These results are in line with the suggestions recommended by (Laufer, 1994; Lin-narud 1986) that there are different roles of lexical sophistication played in spoken and written proficiency.

***Lexical Variation in Undergraduate Students' Research Articles***

Lexical variation of the words employed in the undergraduate students' research articles were identified based on the number of different words, type token ratio, verb diversity, and lexical word diversity. Table 3 contains values which represent the counts of different words found in the research articles. Four measures used are Number of Different Words (NDW), Number of Different Words of first fifty words (NDW-50), Number of Different Words of expected random 50 (NDW-ER50), and Number of Different Words of expected sequence 50 (NDW-ES50).

NDW measure counted number of different words or number of word types in a text. NDW-50 calculated number of different word types in the first fifty words of sample. NDW-ER50 computed the mean of the number of different word of 10 random 50-word samples. NDW-ES50 accounted for the mean of word types of 10 random 50-word sequences. Table 3 consists of number of different word values of the undergraduate students' research articles in four measures. Counting the undergraduate stu-

dents' number of different words in their article using earlier measure such as NDW as well as the transformational measures such as, NDW-50, NDW-ER50, NDW-ES5, all the count results shows that the number of different words of the undergraduate students' articles are higher than those of the values of Chinese oral narratives.

The second way to identify lexical variation of the words employed in the research articles is using type per token ratio (TTR). Table 4 contains values which represent the counts of lexical variation in the research articles using six measures, they are TTR, MSTTR, CTTR, RTTR, LogTTR, and UBER. TTR calculate the number of word types to the number of tokens in the research articles. MSTTR divides a sample into successive segments of a given length and then calculate the average TTR of all segments. MSTTR, CTTR, RTTR, LogTTR, and UBER are TTR transformation with different formula implemented in counting lexical variation in the research articles. The results of TTR analysis using these six measures are found in Table 4. Compared with the type token ratio of Chinese English learners (0.686 & 4.942), the undergraduate students' type token ratio is higher which means the undergraduate students employ number of different word types which are higher than those of Chinese learners.

The third way to analyze lexical variation is by identifying the verb diversity using VV1, SVV1, and CVV1. VV1 counts the ratio of the number of verb types to the total number of verbs in the research articles. Two other transformation of VV1 are SVV1 and CVV1. Both are made to reduce the sample size effect. The results of analysis of the undergraduate students' verb diversity are recorded in the Table 5. The mean values got from these measures are 49.73

**Table 3. The Undergraduate Students' Number of Different Words**

	Lexical Variation Values			
	Number of Different Words (NDW)			
	NDW	NDW-50	NDW-ER50	NDW-ES50
Mean Values	677.2	36.96	38.29	35.82
Maximum Values	1367	47	42.20	41.70
Minimum Values	311	29	33.60	29.80

**Table 4. The Undergraduate Students' Type Token Ratio**

	Lexical Variation Values					
	Type Token Ratio					
	TTR	MSTTR	CTTR	RTTR	LogTTR	UBER
Mean Values	0.18	0.72	7.76	10.98	0.79	17.16
Maximum Values	0.34	0.81	12.26	17.34	0.84	22.28
Minimum Values	0.11	0.65	5.55	7.85	0.74	14.32

**Table 5. The Undergraduate Students’ Verb Diversity**

	Lexical Variation Values		
	Verb Diversity		
	VV1	SVV1	CVV1
Mean Values	0.33	49.73	4.92
Maximum Values	0.63	133.12	8.16
Minimum Values	0.21	19.76	3.14

and 4.92. These values are higher compared with the mean values of Chinese learners (13.415 & 2.556).

Lexical word diversity is one of the indicators of lexical variation. In this present research the lexical word diversity is identified using six measures, they are LV, VV2, NV, Adv.V, and Mod.V. Lexical word variation (LV) calculates the number of word types of lexical word to the total number of lexical word. Verb variation 2 (VV2) counts the number of verb type to the total number of lexical wordss. Noun Variation (NV) accounts for the number of noun type to the total number of lexical words. Adjective Variation (AdjV) reckoned on the number of adjective type to the total number of lexical words. Adverb Variation (AdvV) computes the number of adverb type to the total number of lexical word. Modifier Variation (ModV) calculated the number of adjective and adverb types to the total number of lexical words. The count results of lexical word diversity using six measures: LV, VV2, NV, AdjV, AdvV, and ModV are found in Table 6.

The mean values of lexical word diversity of the undergraduate students’ articles and LV, VV2, NV, AdjV, AdvV, and ModV are 0.28, 0.08, 0.25, 0.06, 0.02, and 0.08. Compared with the mean values of Chinese learners using the same measures, they are 0.573, 0.193, 0.590, 0.108, 0.042, and 0.150. The counts explain that the undergraduate students employ fewer lexical variation, verb variation, noun variation, adjective variation, adverb variation, and modifier variation.

**Syntactic Complexity in Undergraduate Students’ Research Articles**

Syntactic complexity is characterized by varieties and sophistication of the grammatical structures

employed in the sentences. Variation and sophistication of the undergraduate students’ syntactic complexity was identified through the mean length of production unit, sentence complexity, amount of subordination, amount of coordination, and degree of phrasal sophistication (Lu, 2010). Mean length of production units cover mean length of sentences (MLS), mean length of T-units (MLT), and mean length of clauses (MLC), while sentence complexity was identified by counting the ratio of clauses to the number of sentences (C/S). Amount of subordination embraced of dependent clause per clause (DC/C) and dependent clause per T-unit (C/T). Amount of coordination comprises of coordinate phrase per clause (CP/C), coordinate phrase per T-unit (CP/T), and T-units per sentence (CN/T). Degree of phrasal sophistication included complex nominals per clause (CN/C) and complex nominals per T-unit (CN/T).

In the same way with lexical complexity, in order to know how high the complexity of the undergraduate students’ sentences, syntactic complexity values of the undergraduate students’ research articles are compared to the syntactic complexity values of NNS-High Chinese Learners’ argumentative essays in WECCCL (Written English Corpus of Chinese Learners) and of NSs’ argumentative essays in LOCNESS (Louvain Corpus of Native English Essays) (Lu, 2010; Ai & Lu, 2013).

**Mean Length of Production Unit**

Mean length of production unit was defined by the mean of number of words existed in sentence, T-unit, and clause. Table7 shows the mean of number of words in each production unit in the undergraduate students’ articles.

**Table 7. The Undergraduate Students’ Length of Production Unit**

	Length of Production Unit Values		
	MLS	MLT	MLC
Mean Values	20.771	19.279	11.076
Maximum Values	28.716	26.808	14.525
Minimum Values	14.626	14.047	8.245

**Table 6. The Undergraduate Students’ Lexical Word Diversity**

	Lexical Variation Counts					
	Lexical Word Diversity					
	LV	VV2	NV	Adj.V	Adv.V	Mod.V
Mean Values	0.28	0.08	0.25	0.06	0.02	0.08
Maximum Values	0.51	0.14	0.45	0.12	0.04	0.16
Minimum Values	0.17	0.04	0.14	0.03	0.01	0.04

The length of production unit of the undergraduate students' research articles are longer than those of NNS-High in WECCL (Written English Corpus of Chinese Learners) and even longer than NS (essays written by American University students in LOCNESS). Respectively, the mean values of MLS, MLT, and MLC measures are Undergraduate Students' research Articles (20.711; 19.279; 11.76), NNS-High (16.632; 14.815; 9.04), and NS (19.153; 17.072; 9.942). This is different with what is found by Hinkel (2011, p. 529) who reveals that NNS relies on shorter sentences and clauses (T-unit) with fewer words per clause.

**Sentence Complexity in Undergraduate Students' Research Articles**

The second criteria of syntactic complexity is determined by the number of clauses in sentence. The mean value for the sentence complexity of the articles was 1.891. This value means that each sentence consists of at least one clause (see Table 8).

**Table 8. The Undergraduate Students' Sentence Complexity**

	C/S
Mean Values	1.891
Maximum Values	3.209
Minimum Values	1.439

The mean value of the undergraduate students' research articles using C/S is 1.891. While, the mean value of argumentative essays written by Chinese learners level 3 in WECCL is 1.656 (Lu, 2010, p. 490). This shows that the sentence complexity of the undergraduate students' research articles is higher than that of Chinese learners' argumentative essays.

**Amount of Subordinations in Undergraduate Students' Research Articles**

The third criteria of syntactic complexity were identified through the amount of subordinations employed in the research articles. The mean values of amount of subordination determined the number of clauses per T-unit, complex T-units per T-units, dependent clauses per clause, and dependent clauses per T-units. Were 1.725, 0.470, 0.375, and 0.670 (See Table 9).

Compared with the values of Chinese learners level 3 in WECCL, the mean values produced by C/T and CT/T are 1.514 and 0.386 (Lu, 2010, p. 490). Meanwhile, the mean values of the undergraduate students' research articles produced by DC/C and DC/

**Table 9. The Undergraduate Students' Amount of Subordinations**

	C/T	CT/T	DC/C	DC/T
Mean Values	1.752	0.470	0.375	0.670
Maximum Values	2.760	0.739	0.561	1.548
Minimum Values	1.297	0.267	0.247	0.320

T are 0.375 and 0.670. Compared with the mean values produced by NNS-High of the same measures are 0.346 and 0.568 while the mean values of NS are 0.404 and 0.726 (Ai & Lu, 2013, p. 258). Using C/T and CT/T, the amount of subordination of the undergraduate students' research articles are higher than those of Chinese learners. Using DC/C and DC/T, the amount of subordination of the undergraduate students' research articles are higher than those of NNS-High but are lower than those of NS. This is similar with what is found by Hinkel (2005, p. 621; 2011) and Sylva (1993) that NNS tend to use less subordination than NS.

**Amount of Coordinations in Undergraduate Students' Research Articles**

The fourth criteria of syntactic complexity were defined based on the number of coordinate phrases in clause, numbers of coordinate phrases per T-unit, and numbers of T-units per sentence. The mean values got from these measures were 0.299, 0.516, and 1.077 (Table 10). This is also in line with what Hinkel (2005, p. 621) and Sylva (1993) find that NNS tends to use more coordination than NS.

**Table 10. The Undergraduate Students' Amount of Coordinations**

	CP/C	CP/T	T/S
Mean Values	0.299	0.516	1.077
Maximum Values	0.681	1.063	1.270
Minimum Values	0.138	0.249	0.985

**Degree of Phrasal Sophistication in Undergraduate Students' Research Articles**

The fifth criteria of syntactic complexity were defined based on the degree of Phrasal Sophistication, they are number of complex nominals per clause, complex nominals per T-unit, and verb phrases per T-unit. The mean values got from these measures were 1.397, 2.411, and 1.077 (See Table 11). Using CN/C and CN/T, the undergraduate students' degree of complex nominals per clause, complex nominals per T-unit are the highest followed by NS (1.222 and 2.089) and NNS-High (1.064 & 1.669) (Ai & Lu, 2013, p. 258).

**Table 11. The Undergraduate Students’ Degree of Phrasal Sophistication**

	CN/C	CN/T	VP/T
Mean Values	1.397	2.441	2.392
Maximum Values	2.000	4.000	3.836
Minimum Values	0.898	1.475	1.827

Using VP/T, the undergraduate students’ degree of verb phrase per T-unit is higher than Chinese Learners level III (Lu, 2010, p. 490).

**The Quality of Undergraduate Students’ Research Articles**

Based on the four criteria developed, they are flawless language, convincing rethoric, relieving academic insight, and elegant style. The mean value of the quality of the undergraduate students’ article is 3.5. This mean value is categorized as above good, or only 0.5 left to get the highest score or the criteria of great article. Table 12 describes the frequency of the undergraduate students articles with different values.

**Correlation Between The Undergraduate Students’ Lexical and Syntactic Complexities and The Quality of Research Articles**

*Lexical complexity* which covers three multi dimensional features, namely, lexical density, lexical sophistication, and lexical variation characterizes students’ academic texts. Lexical complexity is manifest in second language use (Wolfe-Quintero et. al. 1998). The higher the students’ level of proficiency, the more intensive and extensive the lexical complexity use is in the students’ texts (Laufer & Nation, 1995). Lexical complexity relates to learners’ ability to communicate effectively in both spoken and written (Lu, 2012).

**Table 12. Frequency of the Quality of Undergraduate Students’ Research Articles**

Research Article’s Value	2.00	2.25	2.75	3.00	3.25	3.38	3.5	3.63	3.75	3.88	4
F	1	1	6	21	24	1	24	1	24	2	32

**Table 14. Correlation between Lexical Sophistication and the Quality of Research Articles**

Measure	Code	Pearson Correlation	Sig. 2-tailed
Lexical Sophistication 1	LS1	-.058	.498
Lexical Sophistication 2	LS2	.125	.145
Verb Sophitication 1	VS1	-.038	.656
Verb Sophitication 2	VS2	.045	.600
Corrected Verb Sophitication 1	CVS1	.093	.281

**Correlation between Lexical Density and the Quality of Research Articles**

The values found from this count are correlated with the values of research articles’ quality using pearson product moment correlation. The result shows that the correlation is negative or indirect and the correlation between lexical density and quality of article is very weak or almost no correlation because the Correlation coeficient is nearly zero. Other value that explain no correlation is the value of sig.2-tailed (.830) which is bigger than .01 or .05. This means the negative and very weak correlation is not significant (See Table 13).

**Table 13. Correlation between Lexical Density and the Quality of Research Articles**

Measure	Code	Pearson Correlation	Sig. 2-tailed
Lexical Density	LD	-.018	.830

The values of lexical sophistication taken from the ratio of advanced or relatively unusual words to the total number of lexical words in the research articles were correlated to the values of quality of articles. The Sig 2-tailed and the coeficient correlation explain that there are no correlation between them (Table 14).

Lexical variation is determined by the number of different words employed in the research articles. Table 15 explains that number of different words are positvely and significantly correlated with the quality of articles which means the increase in number of different words is followed by the increase of quality. This correlation, however, is low. Other results which are correlated significantly are the values of TTR, LV, VV1, VV2, NV, AdjV, and ModV. These correlations are negative or indirect which mean the increase number of word types, lexical words, verb variations



**Table 15. The Correlation between Lexical Variation and the Quality of Research Article**

Measure	Code	Pearson Correlation	Sig. 2-tailed
Number of Different Words	NDW	.198*	.021
NDW (first 50 words)	NDW-50	-.084	.331
NDW (expected random 50)	NDW-ER 50	.014	.869
NDW (expected sequence 50)	NDW-ES 50	-.004	.965
Type/Token Ratio	TTR	-.272**	.001
Mean Segmental TTR (50)	MSTTR-50	.009	.931
Corrected TTR	CTTR	.050	.559
Root TTR	RTTR	.050	.559
Bilogarithmic TTR	LogTTR	-.148	.084
Uber Index	Uber	-.019	.830
Lexical word variation	LV	-.239**	.005
Verb Variation-I	VV1	-.257**	.002
Squared VV1	SVV1	.049	.566
Corrected VV1	CVV1	.061	.477
Verb Variation-II	VV2	-.228**	.007
Noun Variation	NV	-.191*	.025
Adjective Variation	AdjV	-.236**	.006
Adverb Variation	AdvV	-.060	.490
Modifier Variation	ModV	-.207*	.015

\* Correlation is significant at the 0.05 level (2-tailed)

\*\* Correlation is significant at the 0.01 level (2-tailed)

**Table 16. The Correlation between Syntactic Complexity and the Quality of Research Article**

Measure	Code	Pearson Correlation	Sig. (2-tailed)
Mean Length of Sentence	MLS	.056	.516
Mean Length of T-units	MLT	.050	.560
Mean Length of Clause	MLC	.072	.400
Sentence Complexity	C/S	.000	.998
Dependent clauses per clause	DC/C	.026	.762
Dependent clauses per T-unit	DC/T	.012	.885
Coordinate phrases per clause	CP/C	.047	.589
Coordinate phrases per T-unit	CP/T	.038	.661
T-units per sentence	T/S	.032	.711
Complex nominals per T-unit	CN/T	-.004	.965
Complex nominals per clause	CN/C	.008	.929

I & II, adjective variation and modifier variation are not followed by the increase of quality of the research articles.

Students' syntactic complexity use is indicated through the length of production unit, amount of subordination, amount of coordination, and degree of phrasal sophistication. These important aspects of sentence using different measures are correlated with the values of research article quality. Table 16 shows that none of syntactic complexity indicators is correlated with the quality of articles.

The length of production units is not significantly correlated with the quality of research articles. As a description, a research article (no. 88) which gets maximum value on quality just obtains above average on mean length of sentence and mean length of T-unit and obtains below average on mean length of

clause. The discordant results between the quality of research article's and the mean length of production unit show that the quality of undergraduate student's articles is not always followed by the intensive use of syntactic complexity, or more specifically, by the intensive use of certain long sentences, T-units and clauses.

**Sentence complexity** which is characterized by number of clauses per sentence has no correlation with the quality of research articles. The highest score (4) got on quality of research article by student no. 13 is not followed by the same level of value from sentence complexity measure which is just above the minimum value. Even though the mean value of sentence complexity of the undergraduate students is higher than that of Chinese learners essays but the quality of the students' research articles is not in line yet by the employment of sentence complexity. This result

supports the finding made by Lu (2010) that sentence complexity has no correlation with writing proficiency.

*The amount of subordination* of the undergraduate students' research articles including amount of clause per T-unit, amount of complex T-unit per T-unit, amount of dependent clause per clause and amount of dependent clause per T-unit have no correlation with the research article quality. The highest score (4) got by research article no. 47 is not along with the values from the whole measures' results of amount of subordination which is just above minimum values. On the other hand, the maximum values got by article no. 78 are not in line with the value of its quality (2) except for the value of complex T-unit per T-unit (CT/T).

*The amount of coordination* of the undergraduate students' research articles comprising of amount of coordinate phrases per clause, coordinate phrases per T-unit, and T-units per sentence have no correlation with the quality of the students' articles. Article no. 33 which gets highest score (4) on quality, the values got from the four measures are above average and below average.

*Degree of phrasal sophistication* of the undergraduate students' research articles including amount of complex nominals per clause, complex nominals per T-unit, and Verb phrase per T-unit have no correlation with the research article quality. Research article no. 85 which obtains highest score on quality of research article just gets above average on degree of phrasal sophistication. On the other hand, research article no. 96 which gets maximum value on complex nominals per clause and per T-unit obtains 3.250 on quality which is below the mean value of the whole research articles. Research article no. 138 which gets maximum value on verb phrase per T-unit obtains 3.00 on quality.

## CONCLUSIONS

This study investigates the use of lexical and syntactic complexities in the undergraduate students' research articles and their correlation to their quality.

The lexical variation based on the number of different words employed, type token ratio and verb diversity are also high (cf. Lu, 2012). However, the lexical word diversities including lexical variation, verb variation, noun variation, adjective variation, adverb variation, and modifier variation are low (cf. Lu, 2012).

The value of length of production unit is high (cf. Lu, 2010; Ai & Lu, 2013). This denied what Hinkel

(2003, 2005, 2011) had found that NNS relied on shorter sentences and clauses (T-unit) with fewer word per clause. The undergraduate sentence complexity is also high (cf. Lu, 2010). The amount of subordination is low (cf. Lu, 2010; Ai & Lu, 2013). This finding is similar to what Hinkel (2003, 2005, 2011) found that NNSs use less subordination. The amount of coordination determined by the mean values of coordinate phrase in clause and in T-unit are high and the amount of coordination determined by the value of the amount of T-units in sentence is low (cf. Lu, 2010; Ai & Lu, 2013). The former is in line with Hinkel (2003, 2005, 2011) who states that NNSs tend to use more coordination than NS, while the latter is not. Degree of phrasal sophistication shown by the degree of complex nominals per clause and per T-unit and degree of verb phrases per T-unit are high (cf. Lu, 2010; Ai & Lu, 2013).

The quality of the research articles got 3.5 which is categorized as above good. This indicates that the quality is nearly categorized as great article or only 0.5 point needed to achieve the criteria of great articles.

Lexical density and lexical variation is not in line with the quality of the undergraduate students' research articles. Lexical variation has no correlation with the quality except for the number of different word which has positive and low correlation. Type token ratio and lexical verb diversity also have no correlation with the quality. Lexical word diversity including lexical variation, verb variation, noun variation, adjective variation, and modifier variation have negative and significant correlation but low. All indicators of syntactic complexity have no correlation with the undergraduate students' articles.

Considering the implementation of lexical and syntactic complexities contribute to the elegant style of an academic text and to the equivalent quality on the employment of lexical and syntactic complexities on advanced academic texts, even though their roles are small enough in the whole quality of a research article, the lecturers are suggested to give more attention to this part in their classes so that more attention will be given by the students on the employment of more lexical word diversities, more subordination, and less coordination in their academic writing as well as lexical and syntactic complexities in general.

## REFERENCES

Afini & Cahyono, B.Y. (2012). The Lexical Richness in Narrative Texts of XI Graders of MAN 3 Malang.

- English Language Education*, 1(1). Retrieved from <http://journal-online.um.ac.id>.
- Ai, H. & Lu, X. (2010). *A Web-based System for Automatic Measurement of Lexical Complexity*. Paper presented at 27th Annual Symposium of the Computer Assisted Language Consortium (CALICO-10). Retrieved from [http://www.personal.psu.edu/hua126/papers/calico10\\_ai\\_lu.pdf](http://www.personal.psu.edu/hua126/papers/calico10_ai_lu.pdf).
- Ai, H. & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students writing. In Nicolas Ballier, Ana Díaz-Negrillo, and Paul Thompson (Eds.). *Automatic Treatment and Analysis of Learner Corpus Data* (pp. 249–264). Amsterdam: John Benjamins.
- Basthomi, Y. (2006). *The Rethoric of Research Article Introductions Written in English by Indonesians* (Unpublished doctoral dissertation). Pascasarjana Universitas Negeri Malang: Malang.
- Basthomi, Y. (2009). Examining Research Spaces in Doctoral Prospectus. *TEFLIN Journal*, 20(2), 140–157. Retrieved from <http://journal-online.um.ac.id>.
- Brown, H.D. (2007)(3<sup>rd</sup> ed.). *Teaching by Principles: an Interactive Approach to Language Pedagogy*. New York: San Fransisco State University. Retrieved from [http://moodle.adaptland.it/pluginfile.php/4930/mod\\_resource/content/1/329WP\\_08\\_71.pdf](http://moodle.adaptland.it/pluginfile.php/4930/mod_resource/content/1/329WP_08_71.pdf).
- Francis, W.S., Romo, L.F. & Gelman, R. (2002). Syntactic Structure, Grammatical Accuracy, and Content in Second-Language Writing: An Analysis of Skill Learning and On-line Processing. *Bilingual Sentence Processing*. Retrieved from [academics.utep.edu/Portals/321/.../Francis/FrancisRomoGelman02.pdf](http://academics.utep.edu/Portals/321/.../Francis/FrancisRomoGelman02.pdf).
- Hinkel, E. (2003). Simplicity without Elegance: Features of Sentences in L2 and L1 Academic Texts. *TESOL Quarterly*, 37, 275–301. Retrieved from <http://www.elihinkel.org/downloads.htm>.
- Hinkel, E. (2005). Analyses of second language (L2) Texts and What can be Learnt from them. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning*, Lawrence Erlbaum Associates, 615–628. Retrieved from <http://www.elihinkel.org/downloads.htm>.
- Hinkel, E. (2011). What Research on Second Language Writing Tells us and What it doesn't. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning*, 2, 523–538. New York: Routledge. Retrieved from <http://www.elihinkel.org/downloads.htm>.
- Larsen-Freeman, D. (2006). The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English. *Applied Linguistics*, 27(4), 590–619. Retrieved from [www.pebc.org/.../Larsen-Freeman-Complexity-Fluncy-Accrcy-in-Oral-Wt...](http://www.pebc.org/.../Larsen-Freeman-Complexity-Fluncy-Accrcy-in-Oral-Wt...)
- Laufer, B. (1994). The Lexical Profile of Second Language Writing: Does it Chang over Time?. *RELC Journal*, 25, 21–33.
- Laufer, B & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3): 307–322.
- Lu, X. (2010). Automatic Analysis of Syntatic Complexity in Second Language Learners. *International Journal of Corpus Linguistics*, 15(4), 474–496. doi: 10.1075/ijcl.15.4.02lu I.
- 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. 2012. The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*, 96(2), 190–208.
- Mukminatn, N. (1997). *The Differences of Students' Writing Achievements across Different Course Level* (Unpublished doctoral dissertation). Pascasarjana Universitas Negeri Malang: Malang.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Naves, T. (2007). *Analytical Measurers of Learners' Written Language*. Barcelona: University of Barcelona, English Department. Retrieved from [www.ub.edu/GRAL/pubs/Celaya&Naves2009Writing.pdf](http://www.ub.edu/GRAL/pubs/Celaya&Naves2009Writing.pdf).
- Šišková, Z. (2012). Lexical Richness in EFL Students' Narratives. *University of Reading: Language Studies Working Papers*. Retrieved from [http://www.readingconnect.net/web/files/english-language-and-literature/elal\\_lswp\\_vol\\_4\\_siskova.pdf](http://www.readingconnect.net/web/files/english-language-and-literature/elal_lswp_vol_4_siskova.pdf).
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second Language Development in Writing: Measurers of Fluency, Accuracy, & Complexity*. Honolulu: University of Hawaii Press.
- Yau, M.S.S. & Belanger, J. (1984). The Influence of Mode on the Syntactic Complexiity of EFL Students at Three Grade Levels. *TESL Canada Journal*, 2(1), 65–77. Retrieved from <http://teslcanadajournal.ca/index.php/tesl/article>.