LEXICAL AND SYNTACTIC COMPLEXITIES IN UNDERGRADUATE STUDENTS' RESEARCH ARTICLES AND THEIR CORRELATIONS TO THEIR QUALITY

DISSERTATION

BY RATNA DEWI NIM 109657627711



STATE UNIVERSITY OF MALANG DOCTORATE PROGRAM IN ENGLISH LANGUAGE TEACHING AUGUST 2014

LEXICAL AND SYNTACTIC COMPLEXITIES IN UNDERGRADUATE STUDENTS' ARTICLES AND THEIR CORRELATIONS TO THEIR QUALITY

DISSERTATION

Presented to State University of Malang in partical fulfillment of the requirement for the degree of Doctor

> by Ratna Dewi NIM 109657627711

STATE UNIVERSITY OF MALANG DOCTORATE PROGRAM IN ENGLISH LANGUAGE TEACHING AUGUST 2014 This is to sertify that the Doctorate dissertation of Ratna Dewi has been approved by the dissertation advisors for further approval by the board of Examiners.

Malang, August 25, 2014 Advisor I, Prof. Moh. Adnan Latief, M.A., Ph.D. NIP 19510304 197903 1 001 Malang, August 25, 2014 Advisor II, Prof. Dr. Yazid Basthomi, M.A. NIP 19720828 199903 1 002 Malang, August 25, 2014 Advisor_III Dr. Arwiyati W. Murdibjono, Dip. TESL, M.Pd. NIP 19501104 197603 2 001

This is to sertify that the Doctorate dissertation of Ratna Dewi has been approved by the Board of Examiners.

Malang, August 25, 2014

Board of Examiners

Prof. Moh. Adnan Latief, M.A., Ph.D.,

Prof. Dr. Yazid Basthomi, M.A.,

1

Dr. Arwiyati W. Murdibjono, Dip. TESL, M.Pd.,

Min

Prof. Bambang Yudi Cahyono, M.Pd., M.A., Ph.D.,

ni l

Prof. Dr. Lies Amin Lestari, M.A., M.Pd.,

11

Prof. Dr. Punaji Setyasari, M.Pd.,

Acknowledged by

Director of Doctorate Program Prof. Dr. J Nyoman S. Degeng, M.Pd. NIP 19580923 198502 1 001 Chair

Member

Member

Member

Member

Member

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Ratna Dewi
NIM	: 109657627711
Jurusan/Program Studi	: Pendidikan Bahasa Inggris
Fakultas/Program	: Pasca Sarjana/Program Doktor

Menyatakan dengan sebenarnya bahwa disertasi yang saya tulis ini benar-benar tulisan saya, dan bukan merupakan plagiasi baik sebagian atau seluruhnya.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa disertasi ini hasil plagiasi, baik sebagian atau seluruhnya, maka saya bersedia menerima sanksi atas perbuatan tersebut sesuai dengan ketentuan yang berlaku.

Malang, 16 Juni 2014 Yang membuat pernyataan,

Ratna Dewi

ABSTRACT

Dewi, Ratna. 2014. Lexical and Syntactic Complexities in Undergraduate Students' Research Articles and their Correlations to their Quality.
Dissertation, Doctorate Program in English Language Teaching, State University of Malang. Advisors: (I) Prof. M. Adnan Latief, M.A., Ph.D., (II) Prof. Dr. Yazid Basthomi, M.A., (III) Dr. Arwijati W. Murdibjono. Dipl. TESL, M.Pd.

Key Words: lexical complexity, syntactic complexity, quality of articles, undergraduate students

This study investigates lexical and syntactic complexities in the undergraduate students' research articles. It aims to find out the trends of lexical and syntactic complexity uses in the undergraduate students' research articles and to find out how the lexical and syntactic complexities correlated with the quality of research articles. Lexical and syntactic complexities are important constructs in L2 teaching and research since they are integral parts of L2 learners' overall development in the target language.

The present research employs a quantitative design through corpus based analysis. The undergraduate students' research articles were the corpus investigated which could be downloaded in http://jurnal-online.um.ac.id. The research articles in pdf. file were converted into word.file, in which pictures, graphs, tables, figures, references, title and subtitles were deleted. After the process of deletion, the research articles which consisted of number of paragraphs were scanned using ABC American Spelling. Then, they were again converted into txt.file. To count the lexical complexities of the research articles, the txt.files were firsly tagged in Stanford POS Tagger then lemmatized using MORPHA. Next, the ouput was taken by LCA as input. The output of LCA used was the count results of lexical density measure (LD), lexical sophistication measures (LS1, LS2, VS1, VS2, CVS1), lexical variation measures (NDW, NDW-50, NDW-ER, NDW-ES, TTR, MSTTR, CTTR, RTTR, AdjV, AdvV, ModV). Whereas to count the syntactic complexity of the research articles, the txt.files were parsed in STANFORD PARSER, the output 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). All the count results of lexical and syntactic complexity measures were correlated with the values of the quality of research articles which were found from the assessment of two raters. Pearson product-moment correlation was used to find out the correlation between the values found from each measure and the values of the research article quality. In order to know the level of the employment lexical complexity, the values of the lexical complexity in the undergraduate students' research articles were compared with the values of lexical complexity of Chinese learners' spoken narratives, while the same intention was also imparted to the values of the syntactic complexity in the undergraduate students' research articles which were compared with the values of syntactic complexity of argumentative essays of NNS-High of Chinese learners in WECCL and the values of agumentative essays of NS in LOCNESS

The findings show that lexical density, lexical sophistication, and lexical variation covering number of different words, type token ratio, verb diversity in undergraduate students' research articles are high (cf. Lu, 2012). On the other hand, lexical variation related to lexical word diversity including lexical variation, verb variation, noun variation, adjective variation, adverb variation and modifier variation in undergraduate students' research articles are low (cf. Lu, 2012). Values of mean lengths of production units of sentences, T-units, and clauses are high. Sentence complexity shown through the number of clauses in sentence is also high (cf. Lu, 2012). The amount of subordinations in undergraduate students' research articles shown by the number of clauses and complex T-units in T-unit are high (cf. Lu, 2010; Ai & Lu, 2013). Whilst the amount of subordinations shown by the number of dependent clauses in clause and in T-unit are high (cf. Lu, 2010; Ai & Lu, 2013). The amount of coordinations of coordinate phrases in clause and T-unit are high but not with the amount of T-units in sentence (cf. Lu, 2010; Ai & Lu, 2013). Degree of phrasal sophistication comprising complex nominals per clause and per T-unit in undergraduate students' research articles is high (cf. Lu, 2010; Ai & Lu, 2013). The quality of the udergraduate students' research articles is categorized as above good or it was only a half point needed to achieve the criteria of great research articles. This research find that there is no correlation between lexical density, lexical sophistication and lexical variation to the quality of undergraduate students' research articles except for lexical variation related to the number of different words employed that are significantly correlated eventhough it was low. Whilst lexical variation indicated by lexical word diversity covering lexical variation, verb variation, noun variation, adjective variation, and modifier variation have negative and significant correlation but low. The no correlation also goes to the whole indicators of syntactic complexity to the quality of research articles.

As mentioned on the findings above, the undergraduate students need to employ more lexical word variation, noun variation, verb variation, adjective variation, adverb variation, and modifier variation. Similar reason is directed to the employment of syntactic complexity in the undergraduate students' research articles, the undergraduate students need to improve the employment of subordination (dependent clause per clause and per T-unit) and to decrease the employment of coordination (coordinate phrases in clause and in T-unit). Since the presence of lexical and syntactic complexities contribute to the elegant style and characterize advanced academic written texts, some indicators of lexical and syntactic complexities which are less or more employed in the undergraduate students' research articles as mentioned above are suggested to be given more attention by the lecturers of Complex English Grammar and Academic Writing in class and by the undergraduate students in writing their academic texts.

ABSTRAK

Dewi, Ratna. 2014. Lexical and Syntactic Complexity in the Undergraduate Students' Research Articles and their Correlations to their Quality.
Disertasi, Pendidikan Bahasa Inggris, Program Doktor, Universitas Negeri Malang. Pembimbing: (I) Prof. M. Adnan Latief, M.A., Ph.D., (II) Prof. Dr. Yazid Basthomi, M.A., (III) Dr. Arwijati W. Murdibjono. Dipl. TESL, M.Pd.

Key Words: kata kompleks, kalimat kompleks, kualitas artikel, mahasiswa S1

Penelitian ini menganalisis kata dan kalimat kompleks yang terdapat dalam artikel mahasiswa S1 Jurusan Pendidikan Bahasa Inggris dan Sastra Inggris, Fakultas Sastra, Universitas Negeri Malang. Penelitian ini bertujuan menemukan kecenderungan penggunaan kata dan kalimat kompleks dalam artikel mahasiswa S1 dan untuk mengetahui korelasi yang terjadi antara penggunaan kata dan kalimat kompleks dengan kualitas artikel. Kata kompleks dan kalimat kompleks merupakan konstruk penting dalam pengajaran dan penelitian bahasa ke dua karena merupakan bagian integral dari keseluruhan perolehan kemampuan bahasa ke dua pembelajar.

Penelitian ini menggunakan disain penelitian kuantitatif melalui analisis kuantitatif terhadap korpus. Korpus yang dianalisis adalah artikel mahasiswa S1 yang bisa di unduh di http://jurnal-online.um.ac.id. Atikel dalam bentuk file pdf dikonversi menjadi file word, kemudian gambar, grafik, tabel, bagan, referensi, judul dan sub judul dihapus. Setelah proses penghapusan, artikel yang hanya berisi sejumlah paragraf discan menggunakan ABC American Spelling. Selanjutnya, artikel tersebut dikonversi menjadi file txt. Untuk menghitung lexical complexity yang terdapat dalam artikel, pertama-tama kata-kata dalam teks berbentuk file txt diidentifikasi berdasarkan part of speech menggunakan Stanford POS Tagger, kemudian dilakukan pengelompokan menggunakan MORPHA. Output dari pengelompokan tersebut menjadi input bagi LCA. Output LCA berupa hasil hitung dari lexical density (LD), lexical sophistication (LS1, LS2, VS1, VS2, CVS1), lexical variation (NDW, NDW-50, NDW-ER, NDW-ES, TTR, MSTTR, CTTR, RTTR, AdjV, AdvV, ModV). Sedangkan, untuk menghitung kalimat kompleks yang terdapat dalam artikel, teks dalam bentuk file txt dikelompokkan berdasarkan struktur grammatikal kalimat menggunakan STANFORD PARSER, output dari parser ini kemudian dihitung menggunakan TREGEX, selanjutnya dianalisis menggunakan L2SCA. Output L2SCA berupa hasil hitung terhadap panjang unit produksi (MLS, MLC, MLT), kompleksitas kalimat (C/S), jumlah subordinasi (C/T, CT/T, DC/C, DC/T), jumlah koordinasi (CP/C, CP/T, T/S), dan tingkat sophistikasi prase (CN/C, CN/T, VP/T). Semua hasil hitung kata dan kalimat kompleks dikaitkan dengan nilai kualitas artikel hasil penelitian yang diperoleh dari hasil penilaian dua rater. Pearson product-Moment Correlation digunakan untuk menganalisis korelasi antara nilai dari alat hitung kata kompleks dan kalimat kompleks dengan nilai kualitas artikel yang diperoleh dari hasil penilaian interrater berdasarkan rubrik. Untuk mengetahui tingkat penggunaan kata kompleks, semua nilai kata kompleks yang terdapat

dalam artikel hasil penelitian dibandingkn dengan nilai kata kompleks yang terdapat dalam narasi lisan pelajar Cina begitu pula dengan nilai kalimat kompleks dari artikel hasil penelitian mahasiswa S1 yang dibanding dengan essay argumentatif pelajar Cina level tinggi dalam WECCL dan nilai essay argumentatif NS yang terdapat di LOCNESS.

Temuan penelitian ini menunjukkan bahwa lexical density, lexical sophistication, and lexical variation meliputi jumlah kata berbeda (number of different words), rasio jenis kata per jumlah kata (TTR), dan kata kerja berbeda (verb diversity) yang terdapat dalam artikel mahasiswa S1 memperoleh nilai tinggi (cf. Lu, 2012). Di lain pihak, variasi leksikal (lexical variation) berupa lexical word diversity yang indikatornya meliputi lexical variation, verb variation, noun variation, adjective variation, adverb variation, dan modifier variation yang terdapat dalam artikel mahasiswa S1 mempeoleh nilai rendah (cf. Lu, 2012). Ratarata nilai panjang unit produksi kalimat, T-unit, dan klausa mahasiswa S1 adalah tinggi. Begitu pula kalimat kompleks yang ditandai oleh jumlah klausa per kalimat memperoleh nilai tinggi (cf. Lu, 2012). Jumlah subordinasi artikel mahasiswa S1 yang ditandai oleh jumlah klausa dan kompleks T-unit per T-unit juga memperoleh nilai tinggi (cf. Lu, 2010; Ai & Lu, 2013). Sementara jumlah subordinasi yang ditandai oleh jumlah dependen klausa dalam klausa dan dalam T-unit juga memperoleh nilai rendah (cf. Lu, 2010; Ai & Lu, 2013). Jumlah frase koordinasi dalam setiap klausa dan dalam setiap T-unit dalam artikel bernilai tinggi tetapi tidak dengan jumlah T-unit dalam setiap kalimat yang bernilai rendah (cf. Lu, 2010; Ai & Lu, 2013). Tingkat sophistikasi frase yang meliputi jumlah kompleks nominals per klausa dan per T-unit serta tingkat sophistikasi frase berupa frase kata kerja per T-unit dalam artikel bernilai tinggi (cf. Lu, 2010; Ai & Lu, 2013). Kualitas artikel mahasiswa S1 dikategorikan di atas baik atau kurang setengah poin lagi untuk dapat mencapai kriteria artikel hasil penelitian terbaik. Penelitian ini juga menemukan bahwa tidak terdapat korelasi antara lexical density, lexical sophistication dan lexical variation terhadap kualitas artikel mahasiswa S1 kecuali lexical variation yang berkaitan dengan penggunaan jumlah kata yang berbeda yang berkorelasi secara signifikan tetapi rendah dengan kualitas artikel. Sementara variasi leksikal yang diindikasikan oleh leksikal word diversity antara lain variasi leksikal, variasi kata kerja, variasi kata benda, variasi kata sifat, serta variasi modifier mempunyai korelasi negatif dan signifikan tetapi rendah terhadap kualitas artikel. Ketiadaan hubungan juga berlaku terhadap setiap indikator kalimat kompleks dan kualitas artikel.

Berdasarkan temuan-temuan tersebut, mahasiswa S1 perlu menggunakan lebih banyak variasi kata, variasi kata benda, variasi kata kerja, variasi kata sifat, variasi kata keterangan, dan variasi modifier. Mahasiswa S1 juga perlu meningkatkan penggunaan subordinasi (jumlah dependen klausa per klausa dan per T-unit) dan mengurangi penggunaan koordinasi (jumlah prase koordinasi per klausa dan per T-unit). Karena keberadaan kata dan kalimat kompleks berkontribusi terhadap gaya penulisan yang baik dan menjadi ciri bagi tulisan akademik, sejumlah indikator kata dan kalimat kompleks yang masih kurang atau berlebihan digunakan dalam artikel mahasiswa S1 disarankan untuk diberi perhatian khusus oleh para dosen mata kuliah "Complex English Grammar" dan "Academic Writing" ketika mengajar di kelas dan oleh mahasiswa S1 dalam menulis teks akademik.

ACKNOWLEDGMENTS

Praise to Allah SWT for the gracious and mercy which always be given to the researcher so that this dissertation can finally be completed.

I would like to express my deepest appreciation and gratitude to Prof. Mohammad Adnan Latief, Ph.D. (my first advisor), Prof. Dr. Yazid Basthomi, M.A. (my second advisor), and Dr. Arwiyati W. Murdibjono, Dip. TESL, M.Pd. (my third advisor) for their enthusiastic encouragement, valuable and constructive suggestions, and useful critiques during the writing of this dissertation. Without their assistance, I might not be able to complete this dissertation.

I am particularly grateful to Effendi Kadarisman, M.A., Ph.D for the valuable and constructive suggestions as the examiner in my proposal seminar. I would also like to extend many thanks to Prof. Bambang Yudi Cahyono, M.Pd., M.A., Ph.D. who has kindly shared his valuable time as the examiner and gives valuable suggestions and corrections for the improvement of this dissertation. Particular thanks are also devoted to Prof. Dr. Lies Amin Lestari, M.A., M.Pd. for her readiness to spend her time to be the examiner and for valuable suggestions and corrections for the improvement of this dissertation.

Very special thanks should be given to Prof. Xiaofei Lu, the LCA and L2SCA developer and Associate Professor in Pennsylvania State University, USA. His willingness to give his time to analyze my data so generously has been very much appreciated.

Sincere thanks are extended to all my lecturers in the Doctorate Program, State University of Malang: Prof. Kasihani Suyanto, M.A., Ph.D., Prof. Dr. Soenardi Djiwandono, Prof. Ali Saukah, M.A., Ph.D., Prof. Dr. Siusana Kweldju,

Х

M.A., Prof. Mohammad Adnan Latief, M.A. Ph.D., Effendi Kadarisman, M.A., Ph.D, Prof. Bambang Yudi Cahyono, M.Pd., M.A., Ph.D.who have widened my insight in learning and teaching and become amazing models as professional teachers, researchers, and writers for me. May Allah SWT bless them All.

I would also like to thank to Dr. Irwan Akib, M.Pd., Rector of Muhammadiyah University of Makassar; Dr. Andi Sukri Syamsuri, M.Hum., Dean of Faculty of Teacher Training and Education; Erwin Akib, S.Pd., M.Pd., Head of English Department, Faculty of Teacher Training and Education, for the financial and moral support given during my study.

My grateful thanks are also extended to Director of Directorate of Higher Education who granted me with BPPS. My grateful thanks also go to Dr. Sintha Tresnadewi, M.Pd. who has the kindness to validate my scoring rubric.

Very special thanks are given to my friends Cayandrawati Sutiono, Sintha Tresnadewi, Meinarni Susilowati, Yudi Setyaningsih, Rita Amalisa, Abdul Muth'im, Ghufran Ferdiant, Fathur Rosyid, Ary S. Ningrum, Rohmani Nur Indah and Esti Djunining, who have been really good and wonderful friends during my study.

Last but not least, I wish to extend my heartfelt appreciation to my late mom, St. Rostinah, my father, Drs. Ibrahim Mannasai, my husband, Drs. Kandacong Melle, M.Pd., my sons, Ahmad Akram & Azwar Anas, and my daughters, Raisa Amini & Aini Salsabiela for their continuous prayers and unlimited tolerance throughout my study.

xi

TABLE	OF CONTENTS
-------	-------------

ABSTRACT			Page i iii
A CKNOWI EDGE	MENT	2	w
TABLE OF CONT	TENTS	5	vii
LIST OF TABLES	LAND		iv
LIST OF TABLES	•••••••••		IA V
LIST OF ADDEND			X :
LIST OF APPEND			X1
LIST OF ABBREV	IATIC	JNS	X11
CHAPTER I	INTRO	ODUCTION	
	1.1 I	Background of the Study	1
	1.2 I	Research Questions	13
	1.3 \$	Significance of the Study	14
	1.4 \$	Scope of the Study	15
	1.5 I	Definition of Key Terms	16
CHAPTER II	REVIE	EW OF LITERATURE	
	2.1 Le	exical and Syntactic Complexity Acquisition	17
	2.2 Le	exical and Syntactic Complexity	18
	2310	wical and Syntactic Complexity Analyzers	21
	2.5 L	a Deliability of Lovicel and Syntactic Complexity Analyzers	21
	2.4 II	re Reliability of Lexical and Syntactic Complexity Analyzers	24
	2.5 III	portance, Aim, and Function of Lexical and	27
	Sy	Intactic Complexity Research	25
	DEGE		
CHAPTER III	RESE	ARCH METHOD	
	3.1 Re	esearch Design	27
	3.2 Sc	ource of Data	29
	3.3 Re	esearch Instrument	29
	3.4 Da	ata Collection	32
	3.5 Da	ata Analysis	35
	3.5.1	Analyzing Lexical Complexity Using LCA	35
	352	Analyzing Syntactic Complexity Using L2SCA	36
	353	Analyzing the Quality of the Undergraduate Students'	50
	5.5.5	Pasagroh Articles	38
	251	A relycing the Completion of the Undergraduate	50
	3.3.4	Analyzing the Correlation of the Undergraduate	
		Students' Lexical and Syntactic Complexity and the	
		Quality of Research Articles	45
CHAPTER IV	RESE	ARCH FINDINGS	
	4.1 Le	exical Complexity Undergraduate Students' Research	
	Α	rticles	48
	4.1.1	Lexical Density in Undergraduate Students' Research	
		Articles	48
	4.1.2	LexicalSophistication in Undergraduate Students'	
		Articles	49
	4.1.3	Lexical Variation in Undergraduate Students' Research	
		Articles	51
	42 4	Syntactic Complexity in Undergraduate Students' Articles	56
	-T.2 K // 2 1	Length of Production Unit in Undergraduate Students'	50
	+.∠.1	Descere Articles	50
	100	Research America in U. 1. (1997)	58
	4.2.2	Sentence Complexity in Undergraduate Students	50
		kesearch Articles	39
	4.2.3	Amount of Subordination in Undergraduate Students'	
		Research Articles	60
	4.2.4	Undergraduate Students' Amount of Coordination	61

	4.2.5 Degree of Phrasal Sophistication in Undergraduate Student	s'
	Research Articles	
	4.3 The Quality of Undergraduate Students' Research Articles	
	4.4 Correlation between the Undergraduate Students'	
	Lexical and Syntactic Complexity and the Quality of	
	Research Articles	65
	441 Correlation between the Students' Lexical Density	
	and the Quality of Research Articles	66
	A 1 1 Correlation between the Students' Lexical Density	
	and the Quality of Research Articles	67
	4.4.1.2 Completion between the Undergraduate Students'	
	4.4.1.2 Contention between the Undergraduate Students	
	A relate	(0)
		62
	4.4.1.3 Correlation between the Undergraduate Students	
	Lexical Variation and the Quality of Research	
	Articles	69
	4.4.2 Correlation between the Undergraduate Students'	
	Syntactic Complexity and the Quality of	
	Research Articles	
	4.4.2.1 Correlation between the Undergraduate Students'	
	Mean Length of Production Unit and the Quality of	
	Research Articles	
	4.4.2.2 Correlation between the Undergraduate Students'	
	Sentence Complexity ang the Quality of	
	Research Articles	
	4.4.2.3 Correlation between the Undergraduate Students'	
	Amount of Subordination and the Ouality of	
	Research Articles	
	4424 Correlation between the Undergraduate Students'	
	Amount of Coordination and the Quality of	
	Research Articles	74
	1125 Correlation between the Undergraduate Students'	
	T.T.2.5 Conclution between the ondergraduate students	
	Possereb Articles	75
CUADTED V	DISCUSSION	
CHAPTER V	DISCUSSION 5.1 The Undergreeducte Students' Levicel Complexity	77
	5.1 The Undergraduate Students Lexical Complexity	
	5.2 The Undergraduate Students Syntactic Complexity	
	5.3 The Quality of the Undergraduate Students	0.6
	Research Articles	
	5.4 Correlation between the Students' Lexical and Syntactic	
	Complexity and the Quality of Research Articles	101
CHAPTER VI C	CONCLUSIONS AND RECOMMENDATIONS	
	6.1 Conclusions	109
	6.2 Recommendations	112
REFERENCES		114
APPENDICES .		119
CURRICULUM	VITAE	181

LIST OF TABLES

Table

The Undergraduate Students' Lexical Density	49
The Undergraduate Stuents' Lexical Sophistication	50
The Undergraduate Students' Number of Different Words	51
The Undergraduate Students' Type Token Ratio	53
The Undergraduate Students' Verb Diversity	54
The Undergraduate Students' Lexical Word Diversity	55
The Undergraduate Students' Grammatical Structure	57
The Undergraduate Students' Length of Production Unit	58
The Undergraduate Students' Sentence Complexity	59
The Undergraduate Students' Amount of Subordinations	60
The Undergraduate Students' Amount of Coordinations	62
The Undergraduate Students' Degree of Phrasal Sophistication	63
Frequency of the quality of the Undergraduate Students' Research	
Articles	65
The 4 M's of the quality of the Undergraduate Students'	
Research Articles	65
Correlation between Lexical Density and the Quality of	
Research Articles	66
The Correlation between Lexical Sophistication and the Quality of	
Research Articles	67
The Correlation between Lexical Variation and the Quality of	
Research Articles	70
The Correlation between Length of Production Unit and the Quality	
of Research Articles	72
The Correlation between Sentence Complexity and the Research	
Article Quality	73
The Correlation between the Amount of Subordination and the	
Research Article Quality	74
The Correlation between the Amount of Coordination and the	
Research Article Quality	75
The Correlation between Degree of Phrasal Sophistication and the	
Research Article Quality	76
	The Undergraduate Students' Lexical Density

LIST OF FIGURES

Figure

3.1	Design of this Study	28
3.2	The Procedure of Getting Data of Lexical and	
	Syntactic Complexity	33
3.3	The Procedure of Getting Data of the Quality of Research	
	Articles	34
3.4	Description of the Correlation between Lexical Complexity	
	and the Research Article Quality	46
3.5	Description of the Correlation between Syntactic Complexity	
	and the Research Article Quality	47

LIST OF APPENDICES

Appendix

1.	Lexical Density Results from LCA	119
2.	Lexical Sophistication Results from LCA	120
3.	Lexical Variation (NDW) Results from LCA	124
4.	Lexical Variation (Type Token Ratio) Results from LCA	128
5.	Lexical Variation (Verb Diversity & Lexical Word Diversity) Results	
	from LCA	132
6.	Syntactic Complexity (MLP & SC) Results from L2SCA	136
7.	Syntactic Complexity (Amount of Sub & Amount of Coord) Results	
	from L2SCA	140
8.	Syntactic Complexity (Degree of Phrasal Sophistication) Results from	
	L2SCA	144
9.	Scoring Rubric in Assessing the Quality of Research Articles	146
10.	Rater 1 and Rater 2 Values	153
11.	Final Value of the quality of Research Articles	157
12.	Results of Lexical Density Correlated with the Values of Research	
	Article Quality	159
13.	Results of Lexical Sophistication Correlated with the Values of	
	Research Article Quality	160
14.	Results of Lexical Variation (NDW) Correlated with	
	the Values of Research Article Quality	164
15.	Results of Lexical Variation (TTR) Correlated with	
	the Values of Article Quality	167
16.	Results of Lexical Variation (VD) Correlated with the Values	
	of Research Article Quality	171
17.	Results of Lexical Variation (LWD) Correlated with the Values	
	of Research Article Quality	173
18.	Results of Syntactic Complexity Correlated with the Values of Research	
	Article Quality	177
19.	Examples of the Undergraduate Students' Research Articles	186
20.	Curriculum Vitae	204

LIST OF ABBREVIATIONS

AdjV	: Adjective Variation
AdvV	: Adverb Variation
C/S	: Clauses per Sentences
C/T	: Clauses per T-unit
CN/C	· Complex Nominals per Clause
CN/T	· Complex Nominals per Catalor
CP/C	: Coordinate Phrases per Clause
CP/T	: Coordinate Phrases per T-unit
CT/T	: Complex T-units per T-unit
CTTR	: Corrected TTR
CVS1	: Corrected Verb Sophistication
CVV1	: Corrected VV1
	: Dependent Clauses per Clause
DC/T	: Dependent Clauses per Clause
	: L2 Syntactic Complexity Analyzer
LZSCA	. L2 Syntactic Complexity Analyzer
	: Lexical Complexity Analyzer
	· Dilogarithmia TTD
LOGIIK	· Dilogarithillic TTK
	Lexical Sophistication 1
	Lexical Sophistication 2
	: Lexical word variation
MLC	: Mean Length of Clause
MLS	: Mean Length of Sentence
MLI	: Mean Length of 1-unit
Mod V	: Modifier Variation
MSTTR-50	: Mean Segmental TTR (50)
NDW	: Number of Different Words
NDW-50	: NDW (first 50 words)
NDW-ER50	: NDW (expected random 50)
NDW-ES50	: NDW (expected sequence 50)
NV	: Noun Variation
RTTR	: Root TTR
SVV1	: Squared VV1
T/S	: T-units per Sentence
TTR	: Type Token Ratio
UBER	: Uber Index
VP/T	: Verb Phrases per T-unit
VS1	: Verb Sophistication 1
VS2	: Verb Sophistication 2
VV1	: Verb Variation-I
VV2	: Verb Variation-II

ABSTRACT

Dewi, Ratna. 2014. Lexical and Syntactic Complexities in Undergraduate Students' Research Articles and their Correlations to their Quality.
Dissertation, Doctorate Program in English Language Teaching, State University of Malang. Advisors: (I) Prof. M. Adnan Latief, M.A., Ph.D., (II) Prof. Dr. Yazid Basthomi, M.A., (III) Dr. Arwijati W. Murdibjono. Dipl. TESL, M.Pd.

Key Words: lexical complexity, syntactic complexity, quality of articles, undergraduate students

This study investigates lexical and syntactic complexities in the undergraduate students' research articles. It aims to find out the trends of lexical and syntactic complexity uses in the undergraduate students' research articles and to find out how the lexical and syntactic complexities correlated with the quality of research articles. Lexical and syntactic complexities are important constructs in L2 teaching and research since they are integral parts of L2 learners' overall development in the target language.

The present research employs a quantitative design through corpus based analysis. The undergraduate students' research articles were the corpus investigated which could be downloaded in http://jurnal-online.um.ac.id. The research articles in pdf. file were converted into word.file, in which pictures, graphs, tables, figures, references, title and subtitles were deleted. After the process of deletion, the research articles which consisted of number of paragraphs were scanned using ABC American Spelling. Then, they were again converted into txt.file. To count the lexical complexities of the research articles, the txt.files were firsly tagged in Stanford POS Tagger then lemmatized using MORPHA. Next, the ouput was taken by LCA as input. The output of LCA used was the count results of lexical density measure (LD), lexical sophistication measures (LS1, LS2, VS1, VS2, CVS1), lexical variation measures (NDW, NDW-50, NDW-ER, NDW-ES, TTR, MSTTR, CTTR, RTTR, AdjV, AdvV, ModV). Whereas to count the syntactic complexity of the research articles, the txt.files were parsed in STANFORD PARSER, the output 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). All the count results of lexical and syntactic complexity measures were correlated with the values of the quality of research articles which were found from the assessment of two raters. Pearson product-moment correlation was used to find out the correlation between the values found from each measure and the values of the research article quality. In order to know the level of the employment lexical complexity, the values of the lexical complexity in the undergraduate students' research articles were compared with the values of lexical complexity of Chinese learners' spoken narratives, while the same intention was also imparted to the values of the syntactic complexity in the undergraduate students' research articles which were compared with the values of syntactic complexity of argumentative essays of NNS-High of Chinese learners in WECCL and the values of agumentative essays of NS in LOCNESS

The findings show that lexical density, lexical sophistication, and lexical variation covering number of different words, type token ratio, verb diversity in undergraduate students' research articles are high (cf. Lu, 2012). On the other hand, lexical variation related to lexical word diversity including lexical variation, verb variation, noun variation, adjective variation, adverb variation and modifier variation in undergraduate students' research articles are low (cf. Lu, 2012). Values of mean lengths of production units of sentences, T-units, and clauses are high. Sentence complexity shown through the number of clauses in sentence is also high (cf. Lu, 2012). The amount of subordinations in undergraduate students' research articles shown by the number of clauses and complex T-units in T-unit are high (cf. Lu, 2010; Ai & Lu, 2013). Whilst the amount of subordinations shown by the number of dependent clauses in clause and in T-unit are high (cf. Lu, 2010; Ai & Lu, 2013). The amount of coordinations of coordinate phrases in clause and T-unit are high but not with the amount of T-units in sentence (cf. Lu, 2010; Ai & Lu, 2013). Degree of phrasal sophistication comprising complex nominals per clause and per T-unit in undergraduate students' research articles is high (cf. Lu, 2010; Ai & Lu, 2013). The quality of the udergraduate students' research articles is categorized as above good or it was only a half point needed to achieve the criteria of great research articles. This research find that there is no correlation between lexical density, lexical sophistication and lexical variation to the quality of undergraduate students' research articles except for lexical variation related to the number of different words employed that are significantly correlated eventhough it was low. Whilst lexical variation indicated by lexical word diversity covering lexical variation, verb variation, noun variation, adjective variation, and modifier variation have negative and significant correlation but low. The no correlation also goes to the whole indicators of syntactic complexity to the quality of research articles.

As mentioned on the findings above, the undergraduate students need to employ more lexical word variation, noun variation, verb variation, adjective variation, adverb variation, and modifier variation. Similar reason is directed to the employment of syntactic complexity in the undergraduate students' research articles, the undergraduate students need to improve the employment of subordination (dependent clause per clause and per T-unit) and to decrease the employment of coordination (coordinate phrases in clause and in T-unit). Since the presence of lexical and syntactic complexities contribute to the elegant style and characterize advanced academic written texts, some indicators of lexical and syntactic complexities which are less or more employed in the undergraduate students' research articles as mentioned above are suggested to be given more attention by the lecturers of Complex English Grammar and Academic Writing in class and by the undergraduate students in writing their academic texts.

ABSTRAK

Dewi, Ratna. 2014. Lexical and Syntactic Complexity in the Undergraduate Students' Research Articles and their Correlations to their Quality. Disertasi, Pendidikan Bahasa Inggris, Program Doktor, Universitas Negeri Malang. Pembimbing: (I) Prof. M. Adnan Latief, M.A., Ph.D., (II) Prof. Dr. Yazid Basthomi, M.A., (III) Dr. Arwijati W. Murdibjono. Dipl. TESL, M.Pd.

Key Words: kata kompleks, kalimat kompleks , kualitas artikel, mahasiswa S1

Penelitian ini menganalisis kata dan kalimat kompleks yang terdapat dalam artikel mahasiswa S1 Jurusan Pendidikan Bahasa Inggris dan Sastra Inggris, Fakultas Sastra, Universitas Negeri Malang. Penelitian ini bertujuan menemukan kecenderungan penggunaan kata dan kalimat kompleks dalam artikel mahasiswa S1 dan untuk mengetahui korelasi yang terjadi antara penggunaan kata dan kalimat kompleks dengan kualitas artikel. Kata kompleks dan kalimat kompleks merupakan konstruk penting dalam pengajaran dan penelitian bahasa ke dua karena merupakan bagian integral dari keseluruhan perolehan kemampuan bahasa ke dua pembelajar.

Penelitian ini menggunakan disain penelitian kuantitatif melalui analisis kuantitatif terhadap korpus. Korpus yang dianalisis adalah artikel mahasiswa S1 yang bisa di unduh di http://jurnal-online.um.ac.id. Atikel dalam bentuk file pdf dikonversi menjadi file word, kemudian gambar, grafik, tabel, bagan, referensi, judul dan sub judul dihapus. Setelah proses penghapusan, artikel yang hanya berisi sejumlah paragraf discan menggunakan ABC American Spelling. Selanjutnya, artikel tersebut dikonversi menjadi file txt. Untuk menghitung lexical complexity yang terdapat dalam artikel, pertama-tama kata-kata dalam teks berbentuk file txt diidentifikasi berdasarkan part of speech menggunakan Stanford POS Tagger, kemudian dilakukan pengelompokan menggunakan MORPHA. Output dari pengelompokan tersebut menjadi input bagi LCA. Output LCA berupa hasil hitung dari lexical density (LD), lexical sophistication (LS1, LS2, VS1, VS2, CVS1), lexical variation (NDW, NDW-50, NDW-ER, NDW-ES, TTR, MSTTR, CTTR, RTTR, AdjV, AdvV, ModV). Sedangkan, untuk menghitung kalimat kompleks yang terdapat dalam artikel, teks dalam bentuk file txt dikelompokkan berdasarkan struktur grammatikal kalimat menggunakan STANFORD PARSER, output dari parser ini kemudian dihitung menggunakan TREGEX, selanjutnya dianalisis menggunakan L2SCA. Output L2SCA berupa hasil hitung terhadap panjang unit produksi (MLS, MLC, MLT), kompleksitas kalimat (C/S), jumlah subordinasi (C/T, CT/T, DC/C, DC/T), jumlah koordinasi (CP/C, CP/T, T/S), dan tingkat sophistikasi prase (CN/C, CN/T, VP/T). Semua hasil hitung kata dan kalimat kompleks dikaitkan dengan nilai kualitas artikel hasil penelitian yang diperoleh dari hasil penilaian dua rater. Pearson product-Moment Correlation digunakan untuk menganalisis korelasi antara nilai dari alat hitung kata kompleks dan kalimat kompleks dengan nilai kualitas artikel yang diperoleh dari hasil penilaian interrater berdasarkan rubrik. Untuk mengetahui tingkat penggunaan kata kompleks, semua nilai kata kompleks yang terdapat dalam artikel hasil penelitian dibandingkn dengan nilai kata kompleks yang terdapat dalam narasi lisan pelajar Cina begitu pula dengan nilai kalimat kompleks dari artikel hasil penelitian mahasiswa S1 yang dibanding dengan essay argumentatif pelajar Cina level tinggi dalam WECCL dan nilai essay argumentatif NS yang terdapat di LOCNESS.

Temuan penelitian ini menunjukkan bahwa lexical density, lexical sophistication, and lexical variation meliputi jumlah kata berbeda (number of different words), rasio jenis kata per jumlah kata (TTR), dan kata kerja berbeda (verb diversity) yang terdapat dalam artikel mahasiswa S1memperoleh nilai tinggi (cf. Lu, 2012). Di lain pihak, variasi leksikal (lexical variation) berupa lexical word diversity yang indikatornya meliputi lexical variation, verb variation, noun variation, adjective variation, adverb variation, dan modifier variation yang terdapat dalam artikel mahasiswa S1 mempeoleh nilai rendah (cf. Lu, 2012). Rata-rata nilai panjang unit produksi kalimat, T-unit, dan klausa mahasiswa S1 adalah tinggi. Begitu pula kalimat kompleks yang ditandai oleh jumlah klausa per kalimat memperoleh nilai tinggi (cf. Lu, 2012). Jumlah subordinasi artikel mahasiswa S1yang ditandai oleh jumlah klausa dan kompleks T-unit per T-unit juga memperoleh nilai tinggi (cf. Lu, 2010; Ai & Lu, 2013). Sementara jumlah subordinasi yang ditandai oleh jumlah dependen klausa dalam klausa dan dalam T-unit juga memperoleh nilai rendah (cf. Lu, 2010; Ai & Lu, 2013). Jumlah frase koordinasi dalam setiap klausa dan dalam setiap Tunit dalam artikel bernilai tinggi tetapi tidak dengan jumlah T-unit dalam setiap kalimat yang bernilai rendah (cf. Lu, 2010; Ai & Lu, 2013). Tingkat sophistikasi frase yang meliputi jumlah kompleks nominals per klausa dan per T-unit serta tingkat sophistikasi frase berupa frase kata kerja per T-unit dalam artikel bernilai tinggi (cf. Lu, 2010; Ai & Lu, 2013). Kualitas artikel mahasiswa S1 dikategorikan di atas baik atau kurang setengah poin lagi untuk dapat mencapai kriteria artikel hasil penelitian terbaik. Penelitian ini juga menemukan bahwa tidak terdapat korelasi antara lexical density, lexical sophistication dan lexical variation terhadap kualitas artikel mahasiswa S1 kecuali lexical variation yang berkaitan dengan penggunaan jumlah kata yang berbeda yang berkorelasi secara signifikan tetapi rendah dengan kualitas artikel. Sementara variasi leksikal yang diindikasikan oleh leksikal word diversity antara lain variasi leksikal, variasi kata kerja, variasi kata benda, variasi kata sifat, serta variasi modifier mempunyai korelasi negatif dan signifikan tetapi rendah terhadap kualitas artikel. Ketiadaan hubungan juga berlaku terhadap setiap indikator kalimat kompleks dan kualitas artikel.

Berdasarkan temuan-temuan tersebut, mahasiswa S1 perlu menggunakan lebih banyak variasi kata, variasi kata benda, variasi kata kerja, variasi kata sifat, variasi kata keterangan, dan variasi modifier. Mahasiswa S1 juga perlu meningkatkan penggunaan subordinasi (jumlah dependen klausa per klausa dan per T-unit) dan mengurangi penggunaan koordinasi (jumlah prase koordinasi per klausa dan per T-unit). Karena keberadaan kata dan kalimat kompleks berkontribusi terhadap gaya penulisan yang baik dan menjadi ciri bagi tulisan akademik, sejumlah indikator kata dan kalimat kompleks yang masih kurang atau berlebihan digunakan dalam artikel mahasiswa S1 disarankan untuk diberi perhatian khusus oleh para dosen mata kuliah "Complex English Grammar" dan "Academic Writing" ketika mengajar di kelas dan oleh mahasiswa S1 dalam menulis teks akademik.

CHAPTER I

INTRODUCTION

The present study investgates the trends of lexical and syntactic complexities found in undergraduate students' research articles and to know their correlation with the quality of their research articles. This chapter presents the background of the study, research problems, purpose of the study, significance of the study, scope of the study, and operational definition of key terms.

1.1 Background of the Study

The requirement set by the Ministry of Education and Culture that undergraduate and graduate students who have completed their study submit published research articles (Akuntono, 2012; Nuh, 2012) are responded by the English Department, Faculty of Letters, State University of Malang with an online journal that can be accessed through <u>http://journal-online.um.ac.id</u>. This web facilitates the undergraduate students with an online journal that accommodates the students' needs in publishing their research articles. In line with the implementation of this policy, some preparation is waiting to be accomplished. One of them is the students need an environment that systematically brings them to the condition where they consciously put into their thought the aspects related to research article publication. As new writers, they have little experience in some aspects that are partly occluded (Swales 1996 in Bromwich, 2008; Basthomi, 2006). Before the students meet with the more occluded requirements, lexical and syntactic complexities are considered as a basic aspect to be aware of. Both

requirements support quality and potential readability of research article writing in general.

Lexical and syntactic complexities characterize academic written texts of advanced writers. The high proficient writers emerge with the more sophisticated vocabulary, that is the University Word List (UWL) and 'not in the list words (Laufer & Nation, 1995: 316). Students of higher proficiency level tended to produce longer and complex sentences (Mukminatin: 1997) and longer clauses and T-units in the forms of complex phrases such as coordinate phrase and complex nominals (Lu, 2010).

Lexical complexity use describes the writer' 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 complexities 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 meaningfully 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 University Word List, or in 'Not in the List' words. Moreover, complex ideas are commonly written in complex vocabularies and sentences in order to accommodate the needs for describing and explaining

specification. Pertaining to the nature of academic text, a writer needs to implement lexical and syntactic complexities.

In short, academic texts are characterized by the extensive and 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 were 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: 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 the UWL and 'not-in-the-lists' words.

Other research related to Test of Written English explains that lexical and syntactic complexities were one of the important constructs because it can gauge the L2 writers' writing scores (Fraser et al.: 1999 in Hinkle: 2003; Francis et al.: 2002). The scores were 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: 276)

stated that the degree of sophistication of text determined by syntactic complexity was identified through the extensive use of subordinate clauses. The words and sentences employed by the writers in their writing described 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.

A research (Laufer & Nation, 1995) that measured lexical richness (lexical variation) supported the findings above. They investigated two different compositions of the same subjects on three different proficiency levels. The writing result of the three subject levels in composition one showed that the first level used more words in the highest frequency level than the level two as well as level two than level three. In composition two, level one had the largest number of the first 1000 words, and level three is the smallest. The use of UWL of the three levels in composition one showed that level one used the smallest compared with level two and three. In composition two, level one used the smallest and level three used the greatest. The three levels were also different with one another in the

use of 'not in the list' words in both compositions. The result showed that the students' experience in writing affects the students' lexical richness in writing.

Some research related to syntactic complexity conducted in L2 contexts revealed that syntactic complexity in academic text corpus increased in number and quality through ages. Yau and Belanger (1984:66) who studied the expository and narrative texts of different level students of a secondary school in Hongkong found that the compositions written by the higher-grade levels were more syntactically complex than the lower ones. This is also in line with what was examined by Mukminatin (1991: 96-98) who stated that in the higher level students' expository composition, compound complex sentences were found. She also stated that the higher the course level, the more complex sentences they produced. It was shown by the number of complex sentences used by the students in writing I-IV, which increased progressively from 275 to 500 words. The ability to construct complex sentences also developed along the consecutive courses. Lu (2010) investigated the significance of fourteen syntactic complexity measurers in differentiating different language proficiency levels of different four-year colleges in China through syntactic complexity analyzer. The conclusion related to this research suggested that students' essays at higher proficiency levels tended to produce longer clauses and T-units because of increased use of complex phrases such as coordinate phrases and complex nominal. In addition, Kitamura (2012) studied how EFL essays of various written skill levels were different in their use of subordinators. The result showed that the three groups of Japanese college students revealed different subordinators and varied their use of subordinators as

their proficiency increased. Furthermore, the frequency was more often and the types of subordinators were larger used by the higher groups than the lower ones.

One different research of diferent time on lexical and syntactic complexities was done by Hinkel (2003, 2005, and 2011) who described lexical and syntactic complexities of L2 writers' academic texts by comparing them with the native writer's text. Hinkel (2005: 621) reviewed Sylva's summary (1993) that L2 writers did repetition of vocabularies and employed simpler sentences. Related to linguistic features, non-native speakers' (NNSs) prose contained fewer syntactically complex constructions, such as subordinate clause, descriptive adjective phrase, and possessives but more coordinators, sentence transitions and pronouns. Compared to L1 writers, L2 writers had a restricted syntactic and stylistic repertoire, as well as severely limited range of accessible lexis in writing.

A decade later after Silva's investigation, Hinkel (2003) examined NSs and NNSs' placement essays written in several universities. The L2 text corpus comprised texts written by speakers of six languages: Arabic, Chinese, Indonesian, Japanese, Korean, and Vietnamese. All the NNSs were advanced and trained L2 writers. The results showed 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 (Hinkel, 2005: 622). NNSs' productive range of grammar and lexis was comparatively small and consists largely of construction, prevalent in spoken discourse as well as high-frequency, and every day vocabulary items (Hinkel,

2003: 297). In other words, after a period of time learning in NS country, the NNSs' lexical and syntactic complexities still did not achieve the demand of complexity of the NSs' vocabulary and sentences in academic texts.

Hinkel (2011: 529) again conducted a review to compare L1 and L2 writers of similar social and educational background. She concluded that comparing to L1, L2 texts exhibited less lexical variety and sophistication; had smaller lexical density, and lexical specificity, and more problems of vocabulary misuses; relied on shorter sentences and clauses (T-units) with fewer words per clause and fewer words (e.g. noun and modifiers) per verb; repeated content words more often (i.e. nouns, verbs, adjectives, and adverbs); used shorter words (fewer words with two or more syllables), more conversational and high frequency words (e.g. good, bad, ask, talk), incorporated fewer modifying and descriptive prepositional phrases, employed less subordination and two to three times more coordination. Furthermore, L2 texts also employed fewer passive constructions; fewer lexical (e.g. adjectives and adverbs) and syntactic modifiers (e.g. subordinate clauses) of sentences, nouns, and verbs; more emotive and private verbs; significantly higher rates of personal pronouns (e.g. I, we, he) and lower rates of impersonal/referential pronouns (e.g. it, this, one); markedly fewer abstract and interpretive nouns, and nominalizations (e.g. rotation, cognition, analysis); fewer adverbial modifiers and adverbial clauses; fewer epistemic and possibility hedges (e.g. apparently, perhaps) and more conversational hedges (sort of, in a way); more conversational intensifiers, emphatics, exaggeratives, and overstatements (e.g. totally, always, huge, for sure); fewer down toners (e.g. almost, hardly); more lexical softening devices (e.g. maybe).

After the sequence of research above, Hinkel (2011: 530) explained the L2 writers skill level above undermined the quality of their formal prose. She further examined that even after several years of language learning, the micro properties of L2 writers' texts concerning the broad range of features were still different significantly from that of novice NS writers. She also stated that even advanced and highly educated L2 writers in English speaking countries had a severely limited lexical and syntactic repertoire compared to their NS peers.

In the Indonesian context, the studies of micro features in written and spoken academic texts analyze the lexical richness (lexical sophistication) of the XI graders of MAN 3 Malang composition, the teacher talk of non-native English teachers in foreign language classroom, and cohesive devices in papers written by English Department Students of State University of Malang. Afini and Cahyono (2012) found that both male and female students used the 2,000 most frequent words repetitively. This explains that their lexical proficiency was considered low since 79.12 % of the word families used were included as high frequency words. High proficiency was determined by the use of low frequency words intensively (Nation, 2001).

Pritomo (2010) explained that teacher talk of oral production in EFL setting also described low proficiency shown by the intensive use of high frequency words and the less and least in uttering Base Words 2 and 3. The words produced frequently were mostly categorized as determiners, personal pronouns, prepositions, conjunctions, auxiliary verbs, and particles that occurred at Base Word 1. Meanwhile, content words that were located at Base Words 2 and 3 were rarely produced. The teachers' oral productions were categorized as fair.

Ulpiati (2010) examined cohesive devices implemented in the students' papers of 1500 words in length as their final project of Comparative Literary Studies course. She explained that the students employed eleven out of sixteen cohesive devices recommended by Haliday & Hasan (1984). Three cohesive devices that were most frequently used were (lexical reiteration (56.1%), pronominal reference (19.1%) and demonstrative reference (18.1%). This means the students did not employ all the cohesive devices recommended by Haliday & Hasan (1984) and the employment of the cohesive devices were not equal in the students' papers.

Apriani (2011) examined Prepositions and Collocation Pattern in Reading Texts of upper Secondary English Textbooks. This research identified three highest collocation patterns indicating time, place, and direction as well as some frequent prepositions that were frequently used in the English books. The result of the research suggested the English teachers in upper secondary school who used the books to give a focus on teaching the prepositions and the collocation patterns identified.

The research conducted in different contexts above exhibits different results. In L1 context, the lexical and syntactic complexities of L2 learners are identified unequal with L1 lexical and syntactic complexities in their academic writing. Through years of learning in L1 context, L2 learner text even is still different significantly from novice L1 texts in regard to the number of features, and so are the advanced L2 with their L1 peers (Sylva, 1993; Hinkel, 2003, 2005, 2011).

In L2 contexts such as Japan (Kitamura, 2012), China (Lu, 2010) research on syntactic complexity has been intensively done. Most research indentified the significant differences exhibited through the complexity of sentences made by different levels in educational institutions. In other words, the more mature the writers, the more intensive lexical and syntactic complexities they employed in their writing.

To the writer's knowledge, in the Indonesian context, in relation to lexical complexity, only a few studies have been done. They are related to lexical richness in the essays done by secondary school students (Fini & Cahyono, 2012), to teacher talk in teaching English (Pritomo, 2012), to prepositions and collocation patterns in English books in junior high school (Apriani, 2011), and to cohesive devices produced in the students' final projects (Ulpiati, 2010). Research on lexical complexity produced by students in academic articles is not established yet.

Meanwhile, the only research related to students' production of syntactic complexity in academic texts was conducted by Mukminatin (1997). The research, in fact, focused on the students' achievement in writing but some of the indicators were the complexity and the increase of sentences produced by the students through consecutive years. A research focuses to the identification of the complexity of sentences produced by Indonesian learners in their academic texts has never been investigated.

Based on the above explanation, this present research was sought to fill out the vacuity of research on lexical and syntactic complexities above. Research articles written by the students, who were in the process of achieving their

undergraduate degree, were the academic texts to be analyzed. The students' research articles were completed by the students of English Department of Faculty of Letters, Malang State University who had passed final examinations for fulfilling their study in the Department. The research articles were the conversion of the student's thesis and the final tasks to be submitted in the forms of research articles which should be published in http://journal-online.um.ac.id. The on-line journal contains 137 research articles, 53 were written by the alumni of English Department, Faculty of Letters, State University of Malang in 2012, while 84 research articles were written by the alumni of 2013.

The research articles made by the undergraduate students were manifestation of their writing competence after learning for years in the English Department. The students had gone through a period of studying in English Department and had been involved in learning activities such as listening, speaking, reading, and writing for the sake of having experience, knowledge and skills in their field that was English language before they came to write the research articles.

Research article is a high level writing which has its own pattern that supports the representation of a research article as an academic text which fulfills the objectiveness of content as well as the process of doing the research which in turn describes the quality of a research article. In general, the components which are compelled their existence in a research article are abstract, introduction, research method, finding, discussion, conclusion and recommendation, as well as references.

Besides the presence of the seven components above, the rethoric employed in each component is considered important in determining the quality of a research article. *An introduction* of a research article establishes a research territory, establishes a niche, and occupies the niche (Basthomi: 2006, 2009; Safnil: 2013; UEFAP). *Method* describes the research design, mentions the sources of data and the instruments, and reporting the procedures of collecting and analyzing data. *Findings* present the results of data analysis and connect the result to the previous research. *Discussion* explains background information, statement of result, (un)expected result, reference to previous research for comparison or support, explanation, exemplification, deduction, and recommendation. *Conclusion* contains direct answer to the proposed questions and fulfillment of the research goal. (Anwar, 2010).

Another criterion which expresses the quality of a research article is the fulfillment of the important points that should be mentioned in each component. For example, it is important to mention the aim and problem, research procedure and result using concise and dense language. Also, it is important that the references mentioned are concorded with the sources mentioned in the content, etc (PPKI: 2010).

As a high level writing also, a writer of a research article is required that basic aspect of writing such as grammatical, puntuation, and spelling errors should not be a problem anymore to the writer. A research article writer is considered as a mature writer in that lexical and syntactic complexities are implemented in a creative way (Larsen-Freeman: 2006; Naves: 2007). In other words, the research article writer has the ability to employ the grammatical

structure of language that a writer tends to expose his/her creativity in producing sophisticated and complex lexis as well as grammatically complex sentenses. These became the reason of these two aspects are involved by the researcher as the quality components of a research article.

Based on the results of reviewing items on lexical and syntactic complexity previously, the number of lexical and syntactic complexities were found significantly different between students in different proficiency levels (Lu: 2010; Mukminatin: 1997; Laufer and Nation: 1995) ; there was a significant difference in the use of lexical and syntactic complexities between NS and NNS (Ai and Lu: 2013; Hinkel: 2003, 2005, 2011); research on lexical and syntactic complexities in academic text in the Indonsian context were never been done. The more specific difference between this research and previous research was this research found out the trends of lexical and syntactic complexities in undergraduate students' research articles using softwares then the results were correlated with the quality of the research articles.

The researber analyzed the trends of lexical complexity by identifying and counting the lexical density, the lexical sophistication, and the lexical variation in the undergraduate students' research articles; and analyzed the syntactic complexity by counting the mean length of production unit, the sentence complexity, the amount of subordination, the amount of coordination, and the degree of phrasal sophistication (Lu: 2010, 2012). The count results got were correlated with the quality of research articles.

1.2 Research Questions

Based on the idea explained above, the present research aimed to find out the answer to the problem that was "how are the lexical and syntactic complexities in the undergraduate students' research articles and their correlation to the quality of the research articles?" This problem is be specifically answered by finding the answers of the following questions:

- 1. How is the lexical complexity in the undergraduate students' research articles?
- 2. How is the syntactic complexity in the undergraduate students' research articles?
- 3. How is the quality of the undergraduate students' research articles?
- 4. How does the lexical and syntactic complexity use correlate with the quality of the undergraduate students' research articles?

1.3 Significance of the Study

The result of this study is expected to contribute theoretical and practical significance to the field of lexical and syntactic complexity acquisition in academic writing.

Theoretically, it is expected to give insight to the body of knowledge of lexical and syntactic complexity acquisition in academic writing. It is expected to enrich the research findings on the employment of lexical and syntactic complexities in academic text especially that used by undergraduate students. In practice, identifying lexical and syntactic features of the undergraduate students' research articles can be used as a base in the expansion of teaching lexical and syntactic complexities employed in academic writing. Also recognizing which lexical and syntactic complexities contributed to the academic nature of the

students' texts could facilitate the development of syllabus, materials and method for teaching and for advising academic writing, they were the use of certain features on lexical and syntactic complexities that were found in undergraduate students' reseach articles.

The result mainly helps the students to prepare themselves with information and knowledge of lexical and syntactic complexities of academic texts written by undergraduate students, which can be valuable as an insight and a reflection in writing academic texts, and to the role of writing exposure during the students' study that contribute to the use of lexical and syntactic complexities as important construct in academic writing.

The research exposes the importance of lexical and syntactic complexities of the students' research articles and the exposure provides a circumstance for the students in preparing themselves with a basic skill, before facing with the aspects of journal writing that is more occluded. Both requirements contribute to the quality and potential readability of a journal writing publication in general.

Editors of a journal also accept information, knowledge and an insight of lexical and syntactic complexities employed in a journal research article in general.

1.4 Scope of the Study

The focus of the study was on **identifying** the trends of lexical complexity reflected in the undergraduate students' research articles vis. lexical density, lexical sophistication, and lexical variation and on **identifying** the syntactic complexity employed in the undergradate students' research articles vis. mean length of production units, sentence complexity, amount of subordination, amount
of coordination and degree of phrasal sophistication, and on **finding** how the correlation between lexical and syntactic complexity use in the undergraduate students' research articles and the quality of undergraduate students' research articles. Lexical and syntactic complexities were analyzed using Lexical Complexity Analayzer (LCA) and L2 Syntactic Complexity Analyzer (L2SCA), while the quality of the undergraduate students' research articles was determined by using the scoring rubric developed.

1.6 Operational Definition of Key Terms

The following operational key terms were defined in order to avoid misunderstandings and misinterpretations.

Lexical complexity referred to the features of language use found in the undergraduate students' research articles covering lexical density, lexical sophistication, and lexical variation.

Syntactic complexity referred to the features of language use found in the undergraduate students' research articles covering mean length of production unit, sentence complexity, amount of subordination, amount of coordination, and degree of phrasal sophistication.

Undergraduate students' research articles refer to the undergraduate students' research articles of the English Department, Faculty of Letters, State University of Malang published on line in <u>http://journal-online.um.ac.id</u>. The research articles were converted from the students' thesis.

Quality of articles was characterized by their flawless language, convincing rethoric, retrieving academic insight, and elegant style.

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter presents review of literatures accommodated to support the idea in implementing this present research covering lexical and syntactic complexity acquisition, lexical and syntactic complexities, lexical and syntactic complexity analyzers.

2.1 Lexical and Syntactic Complexity Acquisition

Production data provide evidence for acquisition (Ellis, 2001:14). A research article produced by a learner is a rich source that depicts competency of language learners not except language competence such as lexical and syntactic complexities. Prove of acquisition is manifested in the employment of lexical and syntactic complexities in the undergraduate students' research articles. The acquisition of lexical complexity is reflected through the implementation of wide array of content words, the intensive use of advanced words, and the extent of different words employed. Syntactic complexity is acquired through the employment of longer sentences, syntactic modifiers or subordinate clauses, and complex phrases.

Furthermore, lexical and syntactic complexities in a text reflect not only that learners have come to advanced language competence because they have implemented the advanced lexical and syntactic indicators above but also have

employ them as a medium for accommodating their ideas in order to communicate appropriately, effectively, and specifically in their writing.

2.2 Lexical and Syntactic Complexities

Most researchers approve the same indicators of lexical and syntactic complexities in written text. Lennon (1990) in Caspi (2010) characterizes lexical and syntactic complexities in writing as a wide array of vocabulary items and syntactic structures. Brown (2001: 305) utilizes mode of complexity by a greater variety of lexical items and usage of longer clauses and more subordination. In line with this, Wolfe-Quintero, Inagaki, and Kim (1998) in Lu (2012: 01) state that second language writing is exhibited by lexical complexity or lexical richness expressed through the sophistication and range of an L2 learners' productive vocabulary while, Foster and Skehan (1996), Ortega (2007), and Wolfe-Quintero (1998) in Lu (2010) view that second language writing contains complex sentences exhibited by the variety and the sophistication of their production units or their grammatical structures.

2.2.1 Lexical Complexity

Ai and Lu (2010) describes lexical complexity as the range and degree of sophistication of L2 learners' productive vocabulary. Further, Read (2000), Ai and Lu (2010), Siskova (2012) establish that the complexity of vocabulary use in academic text covers multi dimensional features, including lexical density, lexical sophistication, and lexical variation. Lexical density exhibits the intensive use of lexical words among the total number of words in a text or the proportion of content words used in the total number of words. Lexical sophistication or lexical

rareness describes the proportion of relatively unusual or advanced words in the learners' text. Lexical variation or also labelled as lexical diversity is the range of a learner's vocabulary as displayed in her or his language use or the type words used.

Lexical density is shown by the ratio of lexical words compared with the total number of words in a text. Lexical words covers nouns, adjectives, verbs (excluding modal verbs and auxiliary verbs such as "be" and "have"), and lexical adverb with adjectival base, including those that can function as both an adjective and an adverb (e.g. fast) and including those formed by attaching the –ly suffix to an adjectival root (e.g., "particularly") (Ai & Lu, 2010: 3-4).

Lexical sophistication or lexical rareness relates to the proportion of relatively unusual or advanced words, or sophisticated lexical words in the learners' texts (Read, 2000: 203; Linnarud, 1986 in Lu, 2010: 04). Lexical sophistication or sophisticated lexical words claimed by Linnarud (1986) as English words introduced at grade nine or later in Swedish educational system, and by Hyltenstam (1988) as words beyond the 7,000 most frequent Swedish words and no significant difference between native and near native Swedish writers (Lu, 2010: 4). Laufer (1994) and Laufer and Nation (1995) considered lexical sophistication as the ratio of the number of sophisticated word type (beyond 2000) to the total number of word types in a text (Wolfe-Quintero et.al, 1998).

Harley and King (1989) proposed sophisticated verbs that are not in the list of 20 or 200 most frequent French verbs to characterize lexical sophistication.

Lu (2010: 4-5) categorized lexical sophistication as the words, lexical words, and verbs not in the list of 2,000 words of the BNC or ANC word list.

Two other labels of lexical variation were lexical diversity (Malveln et.al., 2004; Yu, 2010) and lexical range (Crystal, 1982). Lexical variation refers to the range of vocabulary exhibited in his or her language use (Lu: 2010). The word varieties consisted of words of different bases and of different word types employed by the learners in his or her text. Laufer and Nation (1995) used the term "Lexical Richness" to describe the degree of variety and large of vocabulary of written production. They explained that lexical richness was one variety of factors that affected the overall quality of a piece of writing. Effective use of lexical richness occurred in a well-written composition. Lexical richness of written production is determined through the total number of word types used in the first 1000 most frequent words, in the second 1000 most frequent words, in University Word List, and in the less frequent words.

2.2.2 Syntactic Complexity

Syntactic complexity in the students' written texts (student corpora) is indicated through the categories of the whole sentences performed in the texts. The categories cover the varieties and sophistication of the sentence structure. They are mostly described through the length of unit production of clause, sentence and T-units; the intensive use of subordination, coordination, and range of surface syntactic structure; and degree of sophistication of particular syntactic structures (Ortega, 2003 in Lu, 2010).

The study of syntactic complexity through corpus-based study supported the above criteria. Lu (2010) found that longer clauses and T-units which is

characterized by the increase use of complex phrases such as coordinate phrases and complex nominal tended to produce by higher level proficiency. Kitamura (2012) stated that higher group students employed more often and larger types of subordinators. Hinkel (2011) focused on the criteria of syntactic complexity around the length of sentence and clauses (T-unit), the employment of phrases, subordination, and coordination.

2.3 Lexical and Syntactic Complexity Analyzers

L2 Lexical Complexity Analyzer (L2LCA) is a software program for analyzing lexical complexity in a text and L2 Syntactic Complexity Analyzer (L2SCA) is a software program for analyzing syntactic complexity displayed in a text. Both are developed by Xiofei Lu, a professor from Pennsylvania University, USA.

2.3.1 Lexical Complexity Analyzer (L2LCA)

L2 Lexical Complexity Analyzer (L2LCA) was developed by Lu (2012) by accommodating the results of different researches on lexical complexity measurers in English text. It covers the multidimensional features of language use, namely, lexical density, lexical sophistication, and lexical variation. Lexical density that was originally coined by Ure (1971) measures the ratio of the number of lexical words to total number of words in the text (Lu, 2012). The lexical words defined in this study are nouns, adjectives, verbs (except modal verbs, auxiliary verbs, "be" and "have") and adverbs with an adjectival base (it can function as adverb and adjective such as fast) and those formed by attaching –ly suffix to an adjectival root.

Lexical sophistication or lexical rareness counted the proportion of relatively unusual or advanced words in the learner's text (Read, 2000 in Ai, 2010). In identifying lexical sophistication, five measurers are accommodated. Lexical sophistication 1 (LS1) measured the ratio of the number of sophisticated lexical words to the total number lexical words in a text (Linnarud, 1986; Hyltenstam, 1988). Lexical sophistication 2 (LS2) counted the ratio of the number of sophisticated word types (beyond 2000 words) to the total number of word types in a text (Laufer, 1994; Laufer & Nation, 1995). Verb sophistication-I computed the ratio of the number of sophisticated verb types to the total number of verbs in a text (Harley & King, 1989). Corrected Verb sophistication (VS1) recommended Chaudron and Parker's squared version to reduce the sample size effect (Wolfe-Quintero et.al, 1998). Verb Sophistication –II proposed squared version adapted from Carrol's (1964) (Chaudron & Parker, 1990).

Lexical variation or lexical diversity or lexical range refers to the range of a learner's vocabulary as displayed in his/her language use (Lu, 2012: 05). L2LCA implemented 19 measurers in analyzing lexical variety in a learner's text (Ai, 2010; Lu, 2012). They are Number of Different Words (NDW), first 50 words (NDW-50), expected random (NDW-ER50), expected sequence (NDW-ES50), Type/token ratio (TTR), Mean Segmental TTR (50) (MSTTR-50), Corrected TTR (CTTR), Root TTR (RTTR), Bilogarithmic TTR (BTTR), Uber Index (Uber), D measure (D), Lexical word variation (LV), Verb variation-1 (VV1), Square VV1 (SVV1), Corrected VV1 (CVV1), Verb variation-II (VV2), Noun variation (NV), Adjective variation (AdjV), Adverb variation (AdvV), and Modifier variation (ModV).

2.3.2 Syntactic complexity Analyzer (L2SCA)

L2 syntactic complexity analyzer (L2SCA) is a software system that functions for automatic analyses of syntactic complexity reflected in written language. Since this system is designed for advanced second language proficiency research, the system is developed and evaluated using college-level second language writing data selected from the Written English Corpus of Chinese Learners (WECCL) (Wen et. al. in Lu, 2010: 476-477). In other words, the concordance of the system is the written English Corpus of Chinese Learners (WECCL). This latest system consists of fourteen complexity measurers that have been explored and proposed in the second language development literature, namely, five measures covered in both Wolfe-Quintero et. al. (1998) and Ortega (2003), five other measurers taken because of having been at least one previous study to have at least a weak correlation with or effect for proficiency, and three others were recommended by Wolfe-Quitero et. al. (1998) to pursue further. The measurers are categorized into five types (Lu, 2010: 478):

Three measurers calculate length of production at the clausal, sentential, or T-unit level. They are mean length of clause (MLC), mean length of sentence (MLS), mean length of T-unit (MLT).
 One measure counts a sentence complexity ratio (clauses per sentence, or C/S).
 Four measures reflect the amount of subordinations viz T-unit complexity ratio (clause per T-unit, or C/T), a complex T-unit ratio (Complex T-unit per T-unit, or CT/T), a dependent clause ratio (dependent clauses/clause, DC/C), and dependent clauses per T-unit (DC/T).
 Three measurers calculate the amount of coordination, namely coordinate phrases per clause (CP/C), coordinate phrases per T-unit (CP/T), and a sentence

coordination ratio (T-units/sentence, T/S). 5) Three measures count the relationship between particular syntactic structures and larger production units. They are complex nominals per clause (CN/C), complex nominals per T-unit, and verb phrases per T-unit.

2.4 The Reliability of Lexical and Syntactic Complexity Analyzers

The lexical complexity analyzer took as input similar format texts, whether the text was taken from spoken or written texts. The input was texts which were defined or formatted recognizable by the systems. The LCA was assigned to give a numeric score for each of the 25 measures.

For the sake of measuring the reliability of Lexical Complexity Analyzer (LCA), it was examined the effect of lexical complexity to the quality of oral narratives in 12 (twelve) different groups which were divided into four groups of different proficiencies Levels A, B, C, and D) with the assumption that different proficiency groups employed different number of lexical complexity. A One-way ANOVA indicated significant differences in mean sample length among the four levels. Besides that, it was applied a meta analysis of the results from these groups.

The Syntactic complexity analyzer achieved a high degree of reliability (Lu, 2010: 486). The very strong correlation (ranging from .912 for CT/T to 1.000 for MLS) between the syntactic complexity scores computed by the interannotators who examined the essays proved that the explicit definitions for each indicator provided for the analyzer were responded similarly by the annotators. Moreover, the correlation between the complexity scores computed by the system (L2SCA) and by the annotators for the individual essays were from .834 for CP/C

to 1.000 for MLS. All of the correlatios were significan at the .01 level (Lu, 2010: 485-488).

2.5 Importance, Aim, and Function of Lexical and Syntactic Complexity Research

Academic texts such as research articles written by scholars perform their own characteristics. One of the criteria that characterized academic texts is the intensive and extensive uses of lexical and syntactic complexities. Academic texts are written by writers who have mastered in their field, or who have experienced a period of studying that have brought them to a certain level of mastery. As being scholars or experienced students, they have come to the condition in which fluency and accuracy has not been their interest anymore because they have accustomed to and it has not been a hindrance for them in its implementation (Naves, 2006). They have been in the level where their lexical and syntactic knowledge is used in creative ways that was reflected by the broad range of vocabulary exposure and the variety of syntactic structures in their research articles. The creative use of lexis and syntax is the reflection of linguistic maturity of their writer after a long term of practicing to produce effective written prose (Larsen-Freemam, 2006; Naves, 2007; Hinkel, 2010).

Another reason that might cause the appearance of lexical and syntactic complexities in academic text is the nature of the text itself that loaded complex idea, which needed lexical and syntactic complexities to generate the idea meaningfully. It was a common for a term in one field to convey a specific meaning that was different from 2000 most frequent words.

Differences on the frequency and variety of lexical and syntactic complexities were found in some research done to subjects of dissimilar proficiency or the other way, analysis result on lexical and syntactic complexities of texts could distinguish significantly between low and high groups. Kitamura (2012: 335) found that the diversity of subordinators as indicator of writing proficiency improved significantly among the group of different skill levels.

Research findings on lexical and syntactic complexities described the extent of the students' writing proficiency in terms of lexical and syntactic complexities. (Hinkel, 2010: 535; 2003: 299; 2005: 629) stated that the research findings established information about the students' lexical and syntactic complexity proficiency that could be directed to the choice of instruction that matched with the needs of the students to produce reasonably fluent and accurate texts. The findings can also be used by the researchers to find out the ways to improve students' text production skills to yield more sophisticated syntactic construction and lexis or to attain levels of proficiency necessary to create effective written text.

CHAPTER III

RESEARCH METHOD

This chapter relates to the methodological issues covering research design, source of data, data collection, instrument of getting data and data analysis.

3.1 Research Design

The present research employs a quantitative design through corpus based analysis to examine the trends of the implementation of lexical and syntactic complexities in the undergraduate students' research articles 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. A corpus study is considered fast and accurate to identify patterns in large number of data that human analysis might not notice (Baker, 2011: 111). A pearson product-moment correlation coeficient examines the relationship between two variables that are continous in nature such as score (Salkin, 2000). In other words, Peason Correlation Coeficient measures the strength of linear association between two variables or measures how the two variables correlate (Latief, 2010: 10; 2012: 33). The variables involved have to be measured on either an interval or ratio scale but not needed to be measured on the same scale, such as one variable can be ratio and one can be interval).

The corpus studied is the undergraduate students' research articles published on line at <u>http://jurnal-online.um.ac.id/article/7</u>. Through this corpus study, the trends of the lexical and syntactic complexity uses of each undergraduate student in the research articles could be revealed. The results classify the lexical complexity use based on lexical diversity, lexical sophistication, and lexical variation as well as the syntactic complexity use based on the length of production unit, sentence complexity, amount of subordination, amount of coordination and degree of phrasal sophistication in the research articles.

Lexical density was analyzed using one measure, named LD. Lexical sophistication is analyzed using five measures, they are LS1,LS2,VV1,VV2, and CVS1. Lexical variation was categorized based on number of different words, type token ratio, verb diversity, and lexical word diversity. Number of different words are counted using NDW, NDW-50, NDW-ER50, NDW-ES50. Type token ratio was measured using TTR, MSTTR, CTTR, RTTR, LogTTR, Uber. Verb diversity was analyzed using VV1, SVV1, and CVV1. Lexical word diversity was computed using LV, VV2, NV, AdjV, AdvV, and ModV.



Figure 3.1, Design of this Study

LC	= Lexical Complexity	MLP = Mean Length of production unit
SC	= Syntactic Complexity	SC = Sentence Complexity
LD	= Lexical Density	AS = Amount of Subordination
LS	= Lexical Sophistication	AC = Amount of Coordination
LV	= Lexical Variation	DPS = Degree of Phrasal Sophistication
AQ	= Article Quality	RA = Research Articles

The results of the 25 measures mentioned above are correlated to the quality of research articles examined based on the criteria of a research article using Pearson product-moment correlation (see Fig. 3.2 & 3.3). The result of the correlation is there is or there is no significant correlation between each indicator assigned to reflect lexical and syntactic complexities and the values of the research article qualities. In general, the design of this present research is described in Figure 3.1.

3.2 Sources of Data

The data were taken from <u>http://jurnal-online.um.ac.id/article/7</u>, which up until February 28, 2014 consisted of 137 research articles produced by the undergraduate students of the English Department, Faculty of Letters, State University of Malang.

3.3 Research Instrument

Since the present research is a corpus based analysis, the object to be observed is a collection of research articles written by undergraduate students' of English Department, Faculty of Letters, UM. The instruments of getting data are provided related to the form of text which can be run in the corpus analysis softwares, namely: Stanford POS tagger, MORPHA, lexical complexity analyzer (LCA) and Stanford Parser, Tregex, and syntactic complexity analyzer (L2SCA). The requirement for the data can be run in the analyzers is the data in the form of txt.file. However before the original data which were in the form pdf.file were converted into txt.file, the data should be processed through deleting some contents such as title, student's identity mentioned below the title, sub-headings, graphs, figures, pictures, and references. Deleting these contents was only possibly done in the word.doc.file. The reason for the deletion was the plain text or txt. file could not accommodate pictures, table, graphs, and figures.

Related to the deletion of title, sub-headings, references, L2SCA identify each sentence in the corpus as a group of words which was delimited by one of end punctuation marks. Title, students' identity including names, students' numbers, name of university, and subheadings were a set of words which were not ended with end mark, so the analyzers consider them as a part the next sentence. Furthermore, the reason for the references to be deleted was they followed a specific pattern which was not in line with the pattern of sentence in general.

In addition, the researcher developed a scoring rubric 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 are

flawless language, convincing rethoric, retrieving academic insight, and elegant style. *Flawless language* referred to the absence of grammatical, spelling, and punctuation errors in the research articles. *Convincing rethoric* pertained to the ability of the writer to provide an argument by showing the importance of the topic, exposing how far the topic has been investigated, stating that the topic was extended and continued from previous research, explicitely 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* relates to the fullfillment 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 uses of lexical and syntactic complexities which preserved to a comprehensible meaning.

This scoring rubric was firstly examined and agreed by the advisors. The next step, the scoring rubric was shown to an expert for the purpose of validation. A suggestion was given by the expert to accommodate a qualitative research in the scoring rubric. After the first revision, try out was done to two raters. The first rater suggested making a certain criterion for each score determined. The second rater recommended giving a more explanation on the criteria of elegant style. Then, the researcher made criteria for each component for each value and the criteria for the elegant style component was given more explanation. The result of the second revision was again showed to the expert. The comment from the expert was that the scoring rubric was acceptably implemented but it was too long and

possibly not referred to by the raters. However, the researcher retained the length of the scoring rubric to decrease different perception between the raters and the researcher if the criteria were not specifically mentioned. Additionally, the raters were English Writing lecturers, alumni of Doctorate Program of UM in 2013 and 2014, as well as they had experience and mastery in English writing teaching and research. Besides that, the research articles were examined on each component, so it could decrease the effect of the long scoring rubric. In this case, the researcher asked the raters to examine one component completely, then moved to examining the next component, and so on. The quality of the research article was described by the mean of the four components examined.

After the completion of validation, the research articles were distributed to the 6 raters, three raters became the first raters and three others became the second raters. One other rater or the third writer reexamined the research articles when the score of each component was different more than one point. The score given was 4 for great research article, 3 for good research article, 2 for standard research article, and 1 for bad research article.

3.4 Data Collection

The whole research articles were firstly downloaded from <u>http://jurnal-online.um.ac.id/article/7</u>. The research articles that were analyzed using softwares were processed differently with the ones that were examined by inter raters using the scoring rubric developed. The procedures to process the research articles into data that were ready to be analyzed using softwares are in the following.

The research article texts in the pdf file were converted into word.doc file. Then the word file are left with the paragraphs only without title, subheadings,

figures, graphs, tables, pictures, and references. In other words, the data which is originated from the articles, contained only paragraphs which consist of sentences, T-units, clauses, phrases and words. Next, the data were scanned using ABC American spelling in the computer to determine that the data follow American spelling. This was related to the requirement in analyzing lexical complexity. After the process of scanning using ABC American spelling, the data were then converted into txt.file. The txt. files is the only text type which were able to be analyzed using softwares. Figure 3.4 summarizes the procedure of collecting lexical and syntactic complexity data.

Step 1. The researcher downloaded the undergraduate students' articles from <u>http://jurnal-online.um.ac.id/article/7</u>.

Step 2. The researcher converted the undergraduate students' articles in the form of pdf.file into word.doc file

Step 3. The researcher deleted the title, students' identity, sub-headings, graphs, tables, figures, pictures, and references of the undergraduate students' articles .

Step 4. The researcher scanned the undergraduate students' articles in step 3 with ABC American spelling in the computer.

Step 5. The researcher converted the undergraduate students' articles in step 4 into txt.file.

Figure 3.2 Procedures of Getting Data of Lexical and Sytactic Complexities

The data of the quality of undergraduate students' research articles were taken from the process of scoring done by inter raters. Seven raters were involved in scoring the quality of research articles based on the scoring rubric developed for examining the undergraduate students' research articles. The raters were English lecturers, alumni of Doctorate program in English Language Education of UM in 2013 and 2014. They had experience and mastery in English writing teaching and research. Six raters became first or second raters, and one rater became the third rater who did the reassessment on the component or the whole research articles when the value of each component of the quality of research articles given by rater one and two was different more than one point.

The research articles assessed by the raters were the original form of the research articles in the UM on line jurnal. Next, the values given by the raters for the research articles based on the scoring rubric developed were considered as data which reflected the quality of research articles. The data were then analyzed using frequency statistic.

Step 1. The PDF articles, the scoring rubric, scoring table were distributed to the raters.

Step 2. Each article was assessed by Rater 1 ad Rater 2. The raters assessed the articles to find out the quality of articles based on the scoring rubric develepod.

Step 3. The researcher matched the values of Rater 1 and Rater 2.

Step 4. When there was a difference on the component or the whole of the article quality, Rater 3 did the reassessment on the component or on the whole then decided to choose the value given by Rater 1 or Rater 2.

Step 5. After the reassessment, both values of Rater 1 and Rater 2 were not significantly different, then the researcher chose randomly 5 values from Rater 1 and then 5 values from Rater 2, and so on, except for the values given through the process of reassessment, the values were directly chosen. Step 6. The values got in step 5 became the final scores that reflected the value of each article.

Figure 3.3 The procedures of Getting the Data of the Quality of Research Articles

3.5 Data Analysis

This study was conducted to find out the trends of lexical and syntactic complexities employed in the undergraduate students' research articles and to know how the correlation of the frequency of using lexical and syntactic complexities to the quality of the undergraduate students' research articles. The trends in using lexical and syntactic complexities were analyzed using Lexical Complexity Analyzer (LCA) and Syntactic Complexity Analyzer (L2SCA), while the correlation between the trends of using lexical and syntactic complexities and the quality of the undergraduate students' research articles were analyzed using Pearson product-moment correlation coeficient.

3.5.1 Analyzing Lexical Complexity Using LCA

As it has been mentioned before, the data which were able to be analyzed by LCA were reseach articles which only consist of paragraphs. So title, subheadings, tables, graphics, references, and pictures were deleted from each reseach article. After that, the texts were scanned using ABC American spelling, and then the texts were converted into txt.files. Before the data in the form of plain text format were processed in LCA, the plain texts or the txt.files were

Л

tagged using Stanford POS tagger and lemmatized using MORPHA. These two softwares can be downloaded at http://personal.psu.edu/xxl13/dowload.html.

Next, LCA took the files which had been tagged and lemmatized as INPUT. The input files were organized in the "lemma_tag" format. The output of the L2LCA was a comma-delimited list of 35 field names including 1) a filename field, 2) nine fields for recording counts of sentences, word types, sophisticated word types, lexical word types, sophisticated lexical word types, word tokens, sophisticated word tokens, lexical word tokens, and sophisticated lexical word tokens, and 3) 25 fields for the 25 indices (Lu, 2012). Each of the subsequent lines summarized the results for a specific input file, with a comma-delimited list of 35 values that corresponded to the 35 field names.

Lastly, the output files from L2LCA were loaded into Excel for further statistical analysis. The complete results of the analysis can be seen in Appendix 1, 2, 3, 4, and 5.

3.5.2 Analyzing Syntactic Complexity Using L2SCA

In operating the L2SCA, the procedures to follow in analyzing syntactic complexity in the texts are as follows. The plain text was parsed in STANDFORD PARSER (Klein & Manning, 2003). The output of this syntactic parser was a parsed sample that consisted of a sequence of parse trees, with each parsed tree representing the analysis of the syntactic structure of the sample in the texts.

The next step was to query the parse trees. The system used in this step was TREGEX (Levy & Andrew 2006). Since the sample was tokenized and all tokens, including punctuation marks, wre POS- tagged as part of the parsing process, Tregex (Levy & Andrew 2006) was used to count the number of

occurrences of the other eight units and structures to query the parse tree using a set of manually defined Tregex patterns. Tregex only retrieved those nodes that matched the pattern from the input parse trees. The design of patterns matched the set of production units and syntactic structures used in L2SCA software (Lu, 2010: 480-484). They are:

Sentences, a sentence is a group of words delimited with one of the punctuation marks that signal the end of a sentence: period, question mark, exclamation mark, quotation mark, or ellipsis.

Clauses. A clause is defined as a structure with a subject and a finite verb and includes independent clauses, adjective clauses, adverbial clauses, and nominal clauses.

Dependent Clauses. A dependent clause is a finite adjective, adverbial, or nominal clause

T-units. A T-unit is one main clause plus any subordinate clause or nonclausal structure that is attached to or embedded in it

Complex T-units. A complex T-unit is one that contains a dependent clause. *Coordinate Phrase*. Coordinate phrases counted are adjective, adverb, noun, and verb phrase.

Complex nominals. Complex nominals comprise (i) nouns plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, (ii) nominal clauses, and (iii) gerunds and infinitives in subject position.

Verb phrases. Verb phrases comprises both finite and non-finite verb phrases

The third step is L2 Syntactic Complexity Analyzer (L2SCA) counted the frequency of the 9 structures: words (W), sentences (S), verb phrases (VP),

clauses (C), T-units (T), dependent clauses (DC), complex T-units (CT),

coordinate phrases (CP), and complex nominals (CN). The L2SCA computed the 14 syntactic complexity indices of the text: mean length of sentence (MLS), mean length of T-unit (MLT), mean length of clause (MLC), clauses per sentence (C/S), verb phrases per T-unit (VP/T), clauses per T-unit (C/T), dependent clauses per clause (DC/C), dependent clauses per T-unit (DC/T), T-units per sentence (T/S), complex T-unit ratio (CT/T), coordinate phrases per T-unit (CP/T), coordinate phrases per clause (CP/C), complex nominals per T-unit (CN/T), and complex nominals per clause (CP/C). The output files from L2SCA were loaded to Excel for statistical analysis. The complete output of the L2SCA can be seen in Appendix 6, 7, and 8. Stanford parser and Tregex have been bundled with L2SCA which can be downloaded at http://personal.psu.edu/xxl13/dowload.html.

3.5.3 Analyzing the Quality of the Undergraduate Students' Research Articles

In order to get the same perception in assessing the quality of the undergraduate students' reseach articles, the criteria of great, good, standard, and bad reseach articles were developed in a scoring rubric. The criteria of the reseach articles were determined into four components, they were flawless language, convincing rethoric, retrieving academic insight, and elegant style. Description of the four components are:

1. *Flawless language* was judged to obtain 4 (four) when the reseach article was free from grammatical, punctuation, and spelling errors; to get 3 (three) when the reseach article almost free from grammatical, spelling, and punctuation errors; to get 2 (two) when there were occasional errors but did not represent a

major distraction or obscure meaning; to obtain 1 (one) when the reseach article had some errors that distracted meaning.

2. Convincing rethoric pertained to the ability of the writer to provide an argument. In order to obtain score 4 (four), the criteria to be fullfille were: Background created a research space by establishing a territory (explaining usefulness of the topic investigated, phenomena of the topic investigated, and review of items of previous research), establishing a niche (opposing viewpoint and indicating a gap to previous research, raising questions about previous research, continuing a tradition), and occupying a niche (outlining the purpose or announcing the coverage of the study, announcing the importance of findings, indicating the structure, and evaluating the findings).

Method described the research design (mentioning explicitly or implicitly the type of research design and the acts done to achieve the aim), sources of data (explicitly mentioned and the data taken specified through the research type); instrument (stating the instruments used and describing their function), data collection (describing how the instruments projected to get the information needed) and data analysis (depicting each unit of analysis which reflects each purpose).

Findings consisted of opening paragraph that connect some of the present findings with previous researches. The next paragraphs dealt with the results that are more evaluative and observational. It might also deal with statistical measurers followed by some justificatory commentaries. For qualitative research, result was elaborated under subtopics based on the form, tendency, theme, pattern appeared in the data.

Discussion explained background information, statement of result, (un)expected result, reference to previous research for comparison or support, explanation, exemplification, deduction, and recommendation. In addition to qualitative research, this part also presented the researcher's ideas, relationship between patterns, categories, and position of the outcome to previous theories. *Conclusion* contained direct answer to the proposed research questions and fulfillment of the research goal. For quantitative research, it was ended with the reformulation of the proposed deduction or hypothesis, and supplied by some theoretical or practical suggestion for future researchers. In addition to qualitative research, the main finding, or conclusion should reflect "meaning" of the finding.

The criteria for score 3 were explained in the following.

Background. The writer *established the territory* by explaining the usefulness of the topic, describing the phenomena, reviewing of items of previous research; the writer estalished the niche by raising question about previous research and continuing a tradition; the writer occupied the niche through announcing the coverage of the study, and evaluating the findings. *Method.* The writer described *the research design* (mentioning explicitely the type of research design and explaining the acts to achieve the aims), sources of data (explicitely mentioned and the data taken specified through the research type); *instrument* (stating the instruments used and describing their function); *data collection* (description of how the instruments projected to get the information needed is not proportionally stated); data analysis (explanation on each unit of analysis was in adequate to reflect the unit purpose).

Result. It consisted of opening paragraph explaining the present findings. The next paragraphs dealt with the results that were more observational and evaluative). It might also deal with statistical measurer followed by some justificatory commentaries (quantitative research). For qualitative research, result was elaborated in different paragraph based on the form, tendency, theme, or pattern appeared in the data.

Discussion. It explained background information, statement of result, reference to previous research for comparison or support, explanation, deduction, and recommendations. For qualitative research, this part presented the researcher's ideas, relationship between patterns, catgories, and position of the outcome of previous theories.

Conclusion. It contained direct answer to the proposed research questions and fulfillment of the research goal. For quantitative research, it was ended with the reformulation of the proposed deduction or hyphotesis, and supplied by some theoretical or practical suggestion for future researches. For qualitative research, the main finding or conclusion should reflect "meaning" of the finding.

The criteria for score 2 were as follows:

Background. The writer *established the territory* by stating the usefulness of the topic, describing the phenomena, reviewing concept; the writer estalished the niche through raising question about previous research and continuing a tradition; the writer occupies the niche through announcing the coverage of the study.

Method. The writer described the research design (mentioning explicitly the type of research design and explaining the acts to achieve the aims), sources of data (explicitely mentioned and the data taken specified through the research type); *instrument* (stating the instruments used but not specify their function); data collection (description of how the instruments projected to get the information needed is not proportionally stated); data analysis (explanation on each unit of analysis was inadequate to reflect the unit purpose). *Result.* It consisted of paragraphs that did not completely express the component that specified each of research type. It might also deal with statistical measurers but it was insufficiently elaborated (quantitative research). For qualitative research, result was insufficiently classified and elaborated. Discussion. It explained statement of result which consisted of explanation, deduction, and recommendations. For qualitative research, this part presented the researcher's ideas, relationship between patterns, catgories. *Conclusion.* It contained an artificial answer to the proposed research questions and the fulfillment of the research goal. For quantitative research, it was ended with the restatement of the proposed deduction or hyphotesis, and supplied by some practical suggestions for future researchers. For qualitative research, the main finding or conclusion did not reflect "meaning" of the finding. The criteria for score 1 were determined in the following.

Background. The writer *established the territory* by stating the usefulness of the topic, and reviewing concept; the writer estalished the niche through raising question but replicating a similar tradition ; the writer occupied the niche through announcing the coverage of the study.

Method. The writer described *the research design* (defining explicitely the type of research design but not specifically explained the acts which supported the achievement of the aims and some unrelated explanation mentioned), sources of data (explicitely mentioned and the data taken specified through the research type); *instrument* (stating the instruments used but no explanation on the function); *data collection* (description of how the instruments projected to get the information needed is not stated); data analysis (explanation on each unit of analysis is inadequate to reflect the unit purpose).

Result. It consisted of paragraphs that did not express the component that specified each of research type. It might deal with statistical measurers but it was less elaborated (quantitative research). For qualitative research, result was less classified and less elaborated.

Discussion. It explained statement of result without comparison or support, explanation, deduction, and recommendations to the previous research. For qualitative research, the researcher's ideas were not found, no categorization, no relations between patterns, no categories, and no position of the outcome to previous theories.

Conclusion. It contained a superficial answer to the proposed research questions and the fulfillment of the research goal. For quantitative research, it was ended with the restatement of the proposed deduction or hyphotesis, and supplied by some practical suggestions for future researchers. For qualitative research, the main finding or conclusion did not reflect "meaning" of the finding.

3. *Retrieving academic insight* was referred to the fullfillment of important aspects in each component of a reseach article. In order to get score 4 (four), the criteria which should be totally fulfilled were:

Abstract contained a concise and dense statement of problem and purpose, research procedure and result. In addition to qualitative research, the subjects observed were included.

Background incorporated the rational of the research, the problems and the concepts of ways to solve and the announcement of the research purpose. In addition to qualitative research, focus and concept were explored. *Method* provided ways of collecting and analyzing data. In addition to qualitative research, the presence of the researcher/s, subjects, informants, ways to explore data, setting, and time length of research were also mentioned. Besides that, logical argumentation was required related to legitimacy of the research outcome.

Findings presented results of data analysis and results of hypothesis testing. In addition for qualitative research, it loaded details, which were elaborated under subtopics of the research focus.

Discussion comprised answers of research problems or ways of research aim accomplished, interpretation of findings, integration of findings with previous research. For qualitative research, this part encompassed also the researcher's ideas, the relationship between the research findings and the findings of the previous research.

Conclusion consisted of summary of result and discussion, and recommendations.

References covered the concordance between the literature referred in the reseach article and the content of the references.

For score 3 (three), the criteria above were eighty five percents fulfilled for background, method, findings, and discussion component. For score 2 (two), the criteria above were seventy percents fulfilled for background, method, findings, and discussion component. For score 1 (one), the criteria above were fifty percents fulfilled for background, method, findings, and discussion component.

4. Elegant style was assessed 4 (four) when lexical and syntactic complexities were intensively and extensively used; to get 3 (three) when lexical and syntactic complexities were rather frequently used; to obtain 2 (two) when the lexical and syntactic complexities insufficiently used and their implementation lessened the provision of clarity of the idea expressed; to obtain score 1 (one) when the lexis used was the most frequent words and the syntactic structure used was less various so that they did not specify the ideas expressed.

3.5.4 Analyzing the Correlation between the Undergraduate Students' Lexical and Syntactic Complexity Uses and the Quality Research Articles.

The quantitative count from the LCA and L2SCA produced the trends of the students' lexical and syntactic complexities individually. Lexical Complexity was indicated by lexical density, lexical sophistication, and lexical variation. Lexical density was identified using LD. Lexical sophistication was computed using LS1, LS2, VS1, VS2, and CVS1. Lexical variation was identified based number of different words, type token ratio, verb diversity, and lexical word diversity. Further, number of different words were counted using NDW, NDW- 50, NDW-ER50, and NDW-ES50. Type token ratio was measured using TTR, MSTTR, CTTR, RTTR, LogTTR, Uber. Verb diversity was computed using VV1, SVV1, and CVV1. Lexical word diversity was counted using LV, VV2, NV, AdjV, AdvV, and ModV. All measures produced values which were represented by the frequency count of each indicator of lexical complexity.



Figure 3.4 Description of Correlation of Lexical Complexity with the Research Article Quality

Syntactic complexity of the students' proficiency was described through the mean length of production units, sentence complexity, amount of subordination, amount of coordination, and degree of phrasal sophistication. Mean lenght of production unit was identified using MLS, MLT, and MLC. Sentence complexity was counted using C/S. The amount of subordination was computed using C/T, CT/T, DC/C, and DC/T. The amount of coordination was counted using CP/C, CP/T, and T/S. The last, degree of phrasal sophistication was measured using CN/C, CN/T, and VP/T. All measures produced values which were represented by the frequency count of each indicator of syntactic complexity. Description of measures used to count the indicators of syntactic

complexity and their relation to the quality of undergraduate students' research articles can be seen in Figure 3.5.



Figure 3.5, Description of Correlation of Syntactic Complexity with Research Article Quality

Figure 3.3 reveals that syntactic complexity are indicated by length of production unit, sentence complexity, amount of subordination, amount of coordination, and degree of phrasal sophistication. Degree of complexity of length of prodution unit is computed by three measures, they are MLS, MLT, and MLC. Sentence complexity is measured using C/S. Amount of subordination is counted by DC/C and DC, CP/T, and T/S. Degree of phrasal sophistication is counted CN/C, CN/T, and VP/T. The count results of each measures above are correlated to the values of the artcles' quality.

The last step was the count results of all the measures used to compute the lexical complexity as well as the syntactic complexity were correlated to the values of the quality of research articles using Pearson product-moment correlation (see Figure 3.4 & Figure 3.5).

CHAPTER IV

RESEARCH FINDINGS

The findings in this chapter provide the answers to the research questions in Chapter I. The subtopics cover undergraduate students' lexical complexity, undergraduate students' syntactic complexity, quality of undergraduate students' research articles, and relationship between undergraduate students' lexical and syntactic complexities and the quality of research articles.

4.1 Lexical Complexity in Undergraduate Students' Research Articles

This part is aimed at presenting the answer to the main question asked in the research problem, that is "How is lexical complexity incorporated into 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). To identify the existence of lexical complexity of the undergraduate students' research articles, Lexical Comlexity Analyzer (LCA) is used which resulted in the trends in the employment of lexical density, lexical sophistication, and lexical variation in the students' research articles.

4.1.1 Lexical Density in Undergraduate Students' Research Articles

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 4.1 describes the density of the lexical words identified in the undergraduate students' research articles.

Lexical Density Values
0.52
0.57
0.48

 Tabel 4.1 The Undergraduate Students' Lexical Density

Table 4.1 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 maximum count of the lexical density found in the undergraduate students' research articles is 0.57, while the minimum value of the lexical words used in the research articles is 0.48 of the total words in the research articles.

4.1.2 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. Lexical Complexity Analyzer (LCA) calculated lexical sophistication by accommodating five measures, which include 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 count the ratio of the advanced or unusual words to the total number of lexical words in the research articles. VS1 computes 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.

	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	

Tabel 4.2 The Undergraduate Students' Lexical Sophistication

Tabel 4.2 consists of mean, maximum, and minimum values which represent the result of each measure. Two measures (LS1 & LS2) reveal that the mean value of the students' advanced or relatively unusual words is 0.29 or 0.32 of the total lexical words in the research articles. The maximum value attains 0.47 or 0.48 of the total lexical words, while the minimum value only obtains 0.18 or 0.22 of the total lexical words in the research articles.

The mean value of the advanced or unusual verbs in the research articles measured by VS1 is 0.07 while the ones calculated by VS2 and CVS1 are 2.97 and 1.09 of the total number of lexical verbs in the research articles. The maximum and minimum values of advanced or unusual verbs counted by VS1 are 0.18 and 0.02. On the other hand the ones counted by VS2 are 20.53 and 0.21, while the ones counted by CVS1 are 3.20 and 0.32.

4.1.3 Lexical Variation in Undergraduate Students' Research Articles

Lexical variation of the words employed in the research articles are identified based on the number of different words, type token ratio, verb diversity, and lexical word diversity. Tabel 4.3 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 counts number of different words or number of word types in a text. NDW-50 calculates number of different word types in the first fifty words of sample. NDW-ER50 computes the mean of the number of different word of 10 random 50-word samples. NDW-ES50 accounts for the mean of word types of 10 random 50-word sequences. Table 4.3 consists of lexical variation values of the undergraduate students' research articles in four measures (see Table 4.3).

	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		

Tabel 4.3, The Undergraduate Students' Number of Different Words

The mean, maximum, and minimum values of the undergraduate students' lexical variations in their texts counted using NDW measure are 677.2, 1367, and
311. These values explain that the number of different word types employed by the students whose value is average is 677 different words employed in one research article. Compared with the students who has highest and lowest values of the number of different words in their texts, the maximum and minimum values achieved are 1367 and 311.

The mean, maximum, and minimum values of the undergraduate students' lexical variations in their texts counted using NDW-50 measure are 36.96, 47, and 29. These values explains that the number of different words of the first 50 words sample employed by the students whose value is average is about 36 different words among the first fifty words. Compared with the students who has highest and lowest values of the first 50 word in their texts, the maximum and minimum values achieved are 47 and 29.

The mean, maximum, and minimum values of the undergraduate students' lexical variations in their texts counted using NDW-ER50 measure are 38.29, 42.20, and 33.60. These explain that the mean of the number of different words employed by the students whose value is average is about 38 different words. Compared with the students who has highest and lowest values of 10 random 50 word samples, the maximum and minimum different words implemented are about 42 and 33.

The mean, maximum, and minimum values of the undergraduate students' lexical variations in their texts counted using NDW-ES50 measure are 35.82, 41.70, and 29.80. These values explains that the mean value of the 10 random 50 word sequences implemented by the students whose value is average is about 35 different words. Compared with the students who has highest and lowest values of

the the 10 random 50 word sequences in their texts, the maximum and minimum different words used are about 41 and 29.

The second way to identify lexical variation of the words employed in the research articles is using type per token ratio (TTR). Tabel 4.4 below 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 calcultate the average TTR of all segments. 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 the following table 4.4.

		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

 Tabel 4.4 The Undergraduate Students' Type Token Ratio

Table 4.4 reveals that the mean, maximum, and minimum values of the number of word types compared with the number of tokens employed in the students' research articles using TTR measure are 0.18, 0.34, and 0.11. It also informs that the mean, maximum, and minimum values of the students' lexical

variation counted using MSTTR and LogTTR are matched closely, they are 0.72 - 0.79; 0.81 - 0.84; 0.65 - 0.74. In the other side, the mean, maximum, and minimum values of the other three measures, CTTR, RTTR, and UBER have a big different results with the other previous mesurers.

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 tabel 4.5 below.

		Lexical Variation Value	es			
-	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			

Tabel 4.5 The Undergraduate Students' Verb Diversity

Table 4.5 shows that the mean, maximum, and minimum values of the verb diversity in the students' research articles using the three measures are different significantly. VV1 computes mean, maximum, and minimum values of the verb diversity are 0.33, 0.63, and 0.21. In the other side, SVV1 accounts for

the verb diversity as 49.73, 133.12, and 19.76, while CVV1 produce the count of the verb diversity as 4.92, 8.16, and 3.14.

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) reckons on the number of adjective type to the total number of lexical word. Modifier Variation (ModV) calculates 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 4.6 below.

	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

Tabel 4.6 The Undergraduate Students' Lexical Word Diversity

Table 4.6 notifies that the mean, maximum, and minimum values of lexical variation in the students' research articles are 0.28, 0.51, and 0.17. The

mean, maximum, and minimum values of verb variation in the students' research articles are 0.08, 0.14, and 0.04. The mean, maximum, and minimum values of noun variation in the students' research articles are 0.25, 0.45, and 0.14. The mean, maximum, and minimum values of adjective variation in the students' research articles are 0.06, 0.12, 0.03. The mean, maximum, and minimum values of adverb variation in the students' research articles are 0.02, 0.04, and 0.01. The mean, maximum, and minimum values of modifier variation in the students' research articles are 0.02, 0.04, and 0.01. The mean, maximum, and minimum values of modifier variation in the students' research articles are 0.08, 0.16, and 0.04.

4.2 Syntactic Complexity in Undergraduate Students' Research Articles

This part is arranged to answer the main question asked in the research problem, that is "How is syntactic complexity incorporated into undergraduate students' research articles?". Syntactic complexity is characterized by varieties and sophistication of the grammatical structures employed in the sentences. The varieties and the sophistication of the students' sentences is identified through the intensive and extensive use of 14 indices of syntactic complexity . To analyze the existence of syntactic complexity in the undergraduate students' research articles, Syntactic Complexity Analyzer (L2SCA) is used.

In general, the students' research articles defined as the data of this present research comprise of 137 research articles. The mean length of the research articles consist of about 182 sentences, 182 verb phrases, 343 clauses, 196 T-units, 130 dependent clauses, 91 complex T-units, 100 coordinate phrases, and 469 complex nominals. The more specific description of the research articles are mentioned in table 4.7 below. Table 4.7 below retrieves the mean of each grammatical structure.

	Variety Values of Grammatical Structures								
	W	S	VP	С	Т	DC	СТ	СР	CN
Mean Values	3731.365	182.350	465.343	343.066	196.350	130.489	91.496	100.212	469.715
Maximum Values	7246	431	959	726	434	387	205	337	931
Minimum Values	253	11	35	22	11	11	7	6	44

Tabel 4.7 Description of the Undergraduate Students' Grammatical Structures

Table 4.7 notifies that the mean value of the number of words used in the students' research articles are 3731.365, while the maximum number of words used are 7246, and the minimum number is 253 words. The words are written down into the average amount of 182 sentences, the maximum amount of 431 sentences, and the minimal amount of 11 sentences.

The mean, maximum, and minimum values of verb phrases employed in the research articles are 465.343, 959, and 35. The mean, maximum, and minimum values of clauses implemented in the research articles are 343.066, 726, and 22. The mean, maximum, and minimum values of T-units are 196.350, 434, and 11. The mean, maximum, and minimum values of dependent clauses are 130.489, 387, and 11. The mean, maximum, and minimum values of complex Tunits are 91.496, 205, and 7. The mean, maximum, and minimum values of coordinate phrases are 100.212, 337, and 6. The mean, maximum, and minimum values of complex nominals are 469.715, 931, and 44.

Variation and sophistication of the undergraduate students' syntactic complexity is 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 is identified by counting the ratio of clauses to the number of sentences (C/S). Amount of subordination embraces 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 includes complex nominals per clause (CN/C) and complex nominals per T-unit (CN/T).

4.2.1 Length of Production Unit in Undergraduate Students' Research Articles

The common ways to identify the complexity of written text are by counting the length of sentences, T-units, clauses based on the number of words existed in each index.

	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 4.8 The Undergraduate Students' Length of Production Unit

Table 4.8 above informs that the mean, maximum, and minimum values of the students' sentence length are 20.771, 28.716, and 14.626. This means that the mean of sentence length consists of about 20 words. The longest sentence consists of about 28 words, while the shortest sentence consists of about 14 words.

The mean, maximum, and minimum values of the students' T-unit length are 19.279, 26.808, and 14.047. This means that the mean of the students' T-unit length comprises of about 19 words. The longest T-unit comprises of about 28 words, and the shortest T-unit comprises of about 14 words.

The mean, maximum, and minimum values of the students' clause length are 11.076, 14.525, and 8.245. This means that the mean of the students' clause length contains of about 11 words. The longest clause contains of about 14 words, and the shortest clause contains of about 8 words.

4.2.2 Sentence Complexity in Undergraduate Students' Research Articles

One of the criteria of sintactic complexity is determined by the number of clauses in sentence. To identify complexity of sentences employed in written text, C/S measure is used by counting the mean value of the number of clauses to the number of sentences. Table 4.9 retrieves the sentence complexity measures

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

 Tabel 4.9 The Undergraduate Students' Sentence Complexity

The mean, maximum, and minimum values of the number of clauses in students' sentences are 1.891, 3.209, and 1.439. This means the mean of the number of clauses per sentence in students' research articles is about one clause. The maximum numbers of clauses per sentence in students' research articles are

about three clauses and the minimum number of clause per sentence is about one clause.

4.2.3 Amount of Subordinations in Undergraduate Students'

Other criteria of syntactic complexity are identified through the amounts of subordinations implemented in the research articles. The measures used to count for this purpose is clauses per T-unit, complex T-units per T-units, dependent clauses per clause, and dependent clauses per T-units.

	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

 Table 4.10 The Undergraduate Students' Amount of Subordinations

The mean, maximum, and minimum values of the numbers of clauses per T-unit in students' research articles are 1.752, 2.760, and 1.290. This means in each T-unit in the students' research articles contains the average amount of about one clause. The maximum numbers of clauses in one T-unit are about two, while the minimum number is about one clause per T-unit.

The mean, maximum, and minimum values of the numbers of complex Tunits per T-unit in students' research articles are 0.470, 0.730, and 0.267. This means in each T-unit in the students' research articles contains the average amount of about less than one complex T-unit in one T-unit. The maximum number of complexT-unit per T-unit is about less than one, while the minimum number is one complex T-unit per T-unit. The mean, maximum, and minimum values of the numbers of dependent clauses per clause in students' research articles are 0.375, 0.561, and 0.271. This means in each clause in the students' research articles contains the average amount of about less than one dependent clause per clause. The maximum number of dependent clause in one clause is less than one, while the minimum number is about less than one dependent clause per clause.

The mean, maximum, and minimum values of the numbers of dependent clauses per T-unit in students' research articles are 0.670, 1.548, and 0.320. This means in each T-unit in the students' research articles contains the average amount of about less than one dependent clause. The maximum number of dependent clauses in one T-unit is about two, while the minimum number is about less than one dependent clause per T-unit.

4.2.4 Amount of Coordinations in Undergraduate Students' Research Articles

The next component of syntactic complexity is the intensive use of coordination implemented in the research articles. The amount of coordinations implemented in the research articles are identified by counting numbers of coordinate phrases in clause, numbers of coordinate phrases per T-unit, and numbers of T-units per sententence. Tabel 4.11 informs the mean values of the amount of coordination implemented in clause, in T-unit, and in sentence.

	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

 Tabel 4.11 The Undergraduate Students' Amount of Coordinations

The mean, maximum, and minimum values of the numbers of coordinate phrases per clause in students' research articles are 0.299, 0.681, and 0.138. This means in each clause in the students' research articles contains the average amount of about less than one coordinate phrase. The maximum number of coordinate phrase per clause is about less than one, while the minimum number is less than one coordinate phrase per clause.

The mean, maximum, and minimum values of the number of coordinate phrases per T-unit in students' research articles are 0.516, 1.063, and 0.249. This means in each T-unit in the students' research articles contains the average amount of about less than one coordinate phrase. The maximum number of coordinate phrase per T-unit is about one, while the minimum number is less than one coordinate phrase per T-unit.

The mean, maximum, and minimum values of the numbers of T-units per sentence in students' research articles are 1.077, 1.270, and 0.985. This means in each sentence in the students' research articles contains the average amount of about one T-unit. The maximum number of T-unit per sentence is about one, while the minimum number is one T-unit per sentence.

4.2.5 Degree of Phrasal Sophistication in Undergraduate Students' Research Articles

Degree of phrasal sophistication is difined by computing the number of complex nominals per clause, complex nominals per T-unit, and verb phrases per T-unit. Table 4.12 describes the mean values of complex nominals per clause and per T-unit as well as verb phrase per T-unit.

	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
Minimum Values	0.898	1.475	1.827

 Table 4.12 The Undergraduate Students' Degree of Phrasal Sophistication

The mean, maximum, and minimum values of the numbers of complex nominals per clause in students' research articles are 1.397, 2.000, and 0.898. This means in each clause in the students' research articles contains the average amount of about more than one complex nominal. The maximum number of complex nominals per clause is about two, while the minimum number is less than one complex nominals per clause.

The mean, maximum, and minimum values of the numbers of complex nominals per T-unit in students' research articles are 2.441, 4.000, and 1.475. This means in each T-unit in the students' research articles contains the average amount of about more than two complex nominals. The maximum number of complex nominals per T-unit is about four, while the minimum number is more than one complex nominals per T-unit.

The mean, maximum, and minimum values of the number of verb phrases per T-unit in students' research articles are 2.392, 3.836, and 1.439. This means in each T-unit in the students' research articles contains the average amount of about two verb phrases. The maximum numbers of verb phrases in one T-unit are about three, while the minimum number is about one verb phrase in one T-unit.

4.3 The Quality of Undergraduate Students' Research Articles

The part is set up to answer the main question asked in the research problem, that is "How is the quality of the students' research articles". Quality of the students' research articles are measured based on four criteria, they are flawless language, convincing rethoric, retrieving academic insight, and elegant style. Flawless language refers to the non existing of grammatical, spelling, and punctuation errors. Convincing rethoric belongs to the way the writers organize their idea in each part of research article's writing components, namely background, method, result, discussion, and conclusion. Retrieving academic insight pertains to the existence of important aspects to know in research articles in each research articles' components such as abstract, introduction, method, result, discussion, conclusion and suggestion, and references. Elegant style relates to the implementation of lexical and syntactic complexities in the research articles. Table 4.13 describes the research articles quality reflected by the value got based on the criteria of a research article mentioned in the scoring rubric.

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 4.13 Frequency of the Quality of Undergraduate Students' Research Articles

Table 4.13 informs that 3, out of 137 research articles, each has value, 2.00, 2.25, and 3.88 respectively. Six research research articles get 2.75 and 21 research articles get 3.00. 24 research articles obtain 3.25 and 3.75. 24 research articles obtain 3.5. Two research articles get 3.38 and 1 other get 3.63. The rest, 32 research articles obtain 4.

 Table 4.14 The 4 M's of the Undergraduate Students' Research Articles

Ν	Mean	Median	Min	Max
137	3.493	3.500	2	4

The mean value of the undergraduate students' research article in Table 4.14 is 3.493, whilst the median is 3.500. Since no research article gets 3.493, the median value is taken as the nearest value to the mean value. Based on Table 4.13, the number of research articles which get 3.5 mean values are 24.

4.4 Correlation between the Undergraduate Students' Lexical and Syntactic Complexities and the Quality of Research Articles

This part is provided to answer the fourth question, that is "Is there a correlation between lexical and syntactic complexities and the quality of the research articles". To know whether there is a correlation between lexical and syntactic complexities and the quality of research articles, Pearson product-moment correlation is employed.

4.4.1 Correlation between the Undergraduate Students' Lexical Complexity and the Quality of Research Articles

As it has been explained previously, lexical complexity includes multi dimentional features, they are lexical density, lexical sophistication, and lexical variation. Lexical density is analyzed using one measure, lexical sophistication is identified using five measures, and lexical variation is computed using nineteen measures. Each research article is analyzed using the whole measures to identify the occurence of lexical complexity. Then the result of each measure is analyzed with the value of the undergraduate students' research articles' quality in order to know whether there is a correlation between them. The results of the correlation can be found in the following.

4.4.1.1 Correlation between the Undergraduate Students' Lexical Density and the Quality of Research Articles

Lexical density which is characterized by the ratio of lexical word to the total number of words used in the research article is computed by lexical density measure. The result of the computation produces a value that represent the count of the lexical density in the research articles. The values found from this count are correlated with the values of research articles' quality using pearson product moment correlation. The result of analysis can be found in the following Table 4.14.

 Table 4.15 Correlation between Lexical Density and the Quality of Research Articles

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

Tabel 4.14 provides the result of analysis of Pearson product moment correlation which shown by the value of Correlation coeficient that is -.018. This value means 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.

4.4.1.2 Correlation between the Undergraduate Students' Lexical Sophistication and the Quality of Research Articles

Lexical sophistication is represented by advanced or relatively unusual words. The value of lexical sophistication is taken from the ratio of advanced or relatively unusual words to the total number of lexical words in the research articles. The value produced from this analysis is then correlated with the value of the research article quality in order to find out whether there is a correlation between the existence of lexical density in the research article and the quality of the research article. The result of this correlation analysis is found in Table 4.16.

Maggyma	Cada	Pearso	Sig.
Measure	Code	Correlation	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

Table 4.16 Correlation between Lexical Sophistication and the Quality of Research Articles

Pearson product moment correlation is used to know whether there is a correlation between lexical sophistication employed by the students with the

CVS1

.093

.281

quality of their research articles. Correlation coeficient found from the analysis of correlation of LS1 measure with the value of research article quality is -.058, while the value of its sig.2-tailed is .498. Interpretation of the Correlation coeficient and the value of the the sig.2-tailed is the correlation is negative or indirect and very weak because it is nearly zero and is not significant. This means there is no correlation between the presence of advanced or relatively unusual words and the quality of the research articles. In other words, the negative or indirect correalation of the existence of advanced or relatively unusual words with quality of research articles are not significant.

Correlation coeficient which is the count from the value of LS2 measure and the quality of research article using Pearson product moment is .125, while the sig.2-tailed is .145. The value of Correlation coeficient means the correlation is positive or direct, however it is very weak because it is closely to zero and the value the sig.2-tailed means the correlation is not significant. In other words, there is no correlation between lexical sophistication and the quality of the research article. Even if there is a positive or direct and very weak correlation, this correlation is not significant.

Correlation coeficient of the count result of VS1 with the quality of research article is -.038. This value explains that the correlation of advance sophistication verb is negative or indirect with research article quality but the correlation is very weak because (-.038) is nearly zero. Moreover this correlation is not significant because the sig.2-tailed is .656 is bigger than .01 and .05. So, it can be concluded that there is no correlation between the presence of verb sophistication and the quality of writing.

Correlation coeficient of the count result of VS2 with the quality of research article is .045. This value explains that the correlation of advance sophistication verb is positive or direct with research article quality but the correlation is very weak because .045) is nearly zero. Moreover this correlation is not significant because the sig.2-tailed (.656) is bigger than .01 and .05. So, it can be concluded that there is no correlation between the presence of verb sophistication and the quality of writing.

Correlation coeficient of the count result of CVS1 with the quality of research article is .093. This value explains that the correlation of advance sophistication verb is positive or direct with research article quality but the correlation is very weak because .093) is closely to zero. Moreover this correlation is not significant because the sig.2-tailed (.281) is bigger than .01 and .05. So, it can be concluded that there is no correlation between the presence of verb sophistication and the quality of writing.

4.4.1.3 Correlation between the Undergraduate Students' Lexical Variation and the Quality of the Research Articles

Lexical variation is determined by the number of different words employed in the research articles. To identify lexical variation, nineteen measures are used. Some measures are developed to reduce sample size effect of some other previous measures. The result of each measure is correlated to the value of each student's research article using Peason product moment correlation in order to find out whether there is a correlation between both. The value of each Correlation coeficient and the sig.2-tailed are mentioned in Table 4.16.

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

Table 4.17 The Correlation between Lexical Variation and the Quality of Research Article

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

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

The result of five, out of 19 measures above are correlated significantly with research article quality at the level of .01. Lexical variation which is identified by word type per token ratio (TTR) is correlated significantly to the research article quality. This correlation is negative or indirect. This means the increase number of word type is not followed by the increase quality of the research article. The same significant and negative correlations are also faced by the results of LV, VV1, VV2, AdjV measures. The Correlation coeficient of the four counts define the same significant and indirect correlation which means that the increase number of word type is not followed by the increase quality of the Three of the measures' results above are correlated significantly with research article quality at the level of .05. Lexical complexity which is defined by number of different words (NDW) in the research articles is significantly (.21) and positively or directly (198) correlated with research article quality. This means the increase number of different words used is in line with the increase quality of the research articles. The count results of wo other measures, NV and ModV, are significantly correlated with the research article quality but the correlation is indirect or negative. These mean the increase use of lexical complexity which defined as the increase number of noun, adjective and adverb types is not in line with the increase quality of the research articles.

Four other measures (NDW-50, NDWES-50, LogTTR, Uber) have very weak and negative correlation with quality of research articles. This correlation is not significant. So, lexical variation which are counted using these measures has no correlation with quality of research articles. The last four measures (NDW-ER, MSTTR-50, CTTR, RTTR) have very weak and positive correlation which is not significant. In other words, lexical variation use counted using those measures has no correlation with quality of the research articles.

4.4.2 Correlation between the Udergraduate Students' Syntactic Complexity and the Quality of 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. In the following, the count results of these important aspects of sentence using different measures are correlated with the value of research article quality.

4.4.2.1 Correlation between the Udergraduate Students' Mean Length of Production Units and the Quality of Research Articles

One of the criteria of syntactic complexity is indicated by the length of production unit. Production unit of syntactic complexity covers sentence, T-unit, and clause. The mean length of production unit of each research article is correlated to the quality value of each research article using Pearson productmoment correlation in order to know whether there is a correlation between both.

 Table 4.18 The Correlation between Length of Production Units 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

The correlation coeficient as the count result of mean length of sentence with research article quality using Pearson correlation is .056 and the sig.2-tailed is .516. This explains that the correlation is positive but it is very weak because it is nearly

Zero and it is not significant. This means there is no correlation between mean length of sentence and the quality of research article.

The correlation coeficients as the count result of mean length of T-unit and mean length of clause with research article quality using Pearson correlation are .050 and .072. The sig.2-tailed of both are .560 and .400. Both Correlation coeficients elucidate that the positive correalation is very weak because they are nearly zero. Furthermore the sig.2-tailed interpretes that the correlation is not

significat. This means that mean length of T-unit as well as mean length of clause are not correlated with the quality of research article.

4.4.2.2 Correlation between the Undergraduate Students' Sentence Complexity and the Quality of Research Articles

Sentence complexity is identified based on the number of clauses in each sentence. Then the count result is correlated to the quality of research articles in order to know whether there is a correlation between sentence complexity and the quality of the students' research articles.

Tabel 4.19 The Correlation between Sentence Complexity and the Research Article Quality

Measure	Code	Pearson Correlation	Sig. 2-tailed
Sentence Complexity	C/S	.000	.998

Correlation coeficient of the count results of sentence complexity and the quality of research articles is .000 and the sig.2-tailed is .990. This count explains that there is no correlation and not significant between sentence complexity use and the quality of research articles.

4.4.2.3 Correlation between the Undergraduate Students' Amount of Subordination and the Quality of Research Articles

One other indicator of syntactic complexity is amount of subordination in writing. Amount of subordination is reflected through the count of dependent clause per clause (DC/C) and dependent clause per T-unit (DC/T). In order to find out whether there is a correlation between amount of subordination and research article quality, the count results of both measures, DC/C and DC/T, are correlated

with the values of research articles using Pearson product-moment correlation. The results of these analysis are seen in the following Table 4.19.

	Code	Pearson Correlation	Sig. (2-tailed)
Dependent clauses per clause	DC/C	.026	.762
Dependent clauses per T-unit	DC/T	.012	.885

Table 4.20The Correlation between Amount of Subordination and the
Research Article Quality

The correlation coeficient of amount of dependent clauses per clause correlated to research article quality shows that the positive correlation is very weak since they are closely to zero (.026). Further the sig.2-tailed elucidates that the correlation is not significant (.072 > .01 or .05). The correlation coeficient of amount of dependent clauses per T-unit correlated to research article quality shows that the positive correlation is very weak since they are closely to zero (.012). Further the sig.2-tailed elucidates that the correlation is not significant (.085 > .01 or .05). So, it is concluded that the amount subordination has no correlation with the quality of articels.

4.4.2.4 Correlation between the Undergraduate Students' Amount of Coordination and the Quality of Research Articles

The last two of syntactic complexity indicator is amount of coordination. Amount of coordination in this research is computed using three measures, they are CP/C, CP/T, and T/S. The results of these measures are correlated to the value of the research article quality in order to find out whether there is correlation between amount of coordination and the research article quality. The results of these analysis are appeared in Table 4.19.

Maggura	Code	Pearson	Sig.
Wieasure	Coue	Correlation	(2-tailed)
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

 Table 4.21 Correlation between Amount of Coordination and the Quality Research Article

Correlation coeficient of the count results of CP/C correlated with the value of research article quality is .047. It reflects that the positive correlation is very weak since it is nearly zero (.047). The sig.2-tailed (.589 > .01 or .05) interpretes that the positive correlation is not significant. Correlation coeficient of the count results of CP/T correlated with the value of research article quality is .038. It reflects that the positive correlation is very weak since it is nearly zero (.038). The sig.2-tailed (.661 > .01 or .05) interpretes that the positive correlation is not significant. Correlated with the value of T/S correlated with the value of research article quality is .038. It reflects that the positive coefficient of the count results of T/S correlated with the value of research article quality is .032. It reflects that the positive correlation is not significant. Correlated with the value of research article quality is .032. It reflects that the positive correlation is very weak since it is nearly zero (.032). The sig.2-tailed (.711 > .01 or .05) interpretes that the positive correlation is not significant. From the above explanation, it can be concluded that the amount of coordination reflecting through the count results of coordinate phrase per clause, coordinate phrase per T-unit, and T-unit per sentence have no correlation with research article quality.

4.4.2.5 Correlation between the Undergraduate Students' Degree of Phrasal sophistication and the Quality of Research Articles

The last indicator of syntactic complexity is degree of phrasal sophistication. Degree of phrasal sophistication is characterized through the count results of complex nominals per T-unit and complex nominals per clause. The count results of these two measures are correlated with the value of research article quality using Pearson product-moment correlation in order to know whether there is a correlation between degree of phrasal sophistication and research article quality.

Table 4.22	The	Correlation	between	Degree	of	Phrasal	Sophistication	and
	the F	Research Art	icle Quali	ity				

Measure	Code	Pearson Correlation	Sig. (2-tailed)
Complex nominals per T-unit	CN/T	004	.965
Complex nominals per clause	CN/C	.008	.929

The correlation coeficient of complex nominals per T-unit correlated with research article quality is -.004. The sig.2-tailed is .965. This interpretes that complex nominals per T-unit has negative correlation with research article quality but this negative correlation is not significant. So, there is no correlation between complex nominals per T-unit and research article quality. The correlation coeficient of complex nominals per T-unit correlated with research article quality is .008. The sig.2-tailed is .929. This interpretes that complex nominals per clause has positive correlation with research article quality but this positive correlation is not significant. So, there is no correlation is not significant. So, there is no correlation between complex nominals per T-unit and research article quality but this positive correlation is not significant. So, there is no correlation between complex nominals per T-unit and research article quality but this positive correlation is not significant. So, there is no correlation between complex nominals per T-unit and research article quality. The concluded that degree of phrasal sophistication has no correlation with research article quality.

CHAPTER V

DISCUSSION

This chapter provides explanation for the results of analysis that exist in chapter IV and presents the empirical data for the answer of research questions. The discussion mainly focus on the undergraduate students' lexical and syntactic complexities, the quality of the udergraduate students' research articles, and the relationship between the undergradulate students' lexical and syntactic complexities and the quality of research articles.

5.1 The Undergraduate Students' Lexical Complexity

Lexical Complexity Analyzer (LCA) is design by Prof. Xiaofei Lu from Pennsyvania 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 researher 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.

As mentioned in the previous chapters, lexical complexity covers multi dimensional features, namely, lexical density, lexical sophistication, and lexical variation (Read, 2000; Johansson, 2008; Ai & Lu, 2010, and Siskova, 2012). The discussion of findings related to lexical complexity that was explained in Chapter IV following the three features respectively. Since the count results of the softwares are just in the form of scores or numbers, the lexical complexity values of the undergraduate students' research articles are compared to the values of lexical complexity of Chinese learners' spoken naratives in SECCL (Spoken English Corpus of Chinese Learners) in order to know how high the complexity of lexis employed by the undergraduate students are. Some fragments of research articles are displayed to show the employment of lexical complexity empirically.

Lexical Density. As stated in Chapter IV, lexical density is the ratio of the number of lexical words employed to the whole number of words. To see the employment of lexical density in those research articles, some sentences are taken out from research article no. 110 which gets maximum value (0.57) and research article no. 138 which obtaines the minimum value (0.48) as in the following.

In *line* with the *literacy level set* by the *government*, the *focus* of *teaching reading* for *students* of *senior high schools* is to *train students* to *get* the *meaning* of the *texts*. In *other words, reading comprehension* is the *main concern* for them. In a *process* of *teaching* and *learning reading comprehension, questions* may *play* an *important role* (research article no. 110).

The Age of Innocence by Edith Wharton has attracted many experts to make a study of it. For so many years, Wharton's The Age of Innocence has been analyzed through the theory of existentialism, like Cain (2008) and Witherow (2003) do, and through the western point of view in judging every single part of it (research article no. 138).

The bold and italic words in these quotations consist of nouns, adjectives, and verbs (excluding modal verbs, auciliary verbs, "be" & "have") describes the lexical words employed. The appearance of the bold and italic words in the two research articles above expresses the difference in number of the lexical words employed in the two research articles which get maximum and minimum values.

Furthermore 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 is identified through the proportion of the whole lexical sophistication and the whole verb sophistication to the whole lexical words and to the whole lexical verbs. Different formula used in each measure is

an effort to decrease the sample size effect since the longer the text, the higher the repetition of the same sophistication words happens in the texts (Lu, 2012).

The measures used to count lexical sophistcation in this present research are LS1, LS2, VS1, VS2, and CVS1 which compute the ratio of the number of advanced or relatively unusual words to the number of lexical words in a text. As a description in the employment of lexical sophistication in a research article, one paragraph is taken from the research article no. 93 which gets maximum value (0.48) and one from the research article no. 79 which gets minimum value (0.22) using LS2.

Women's social conditions change from time to time. It is generally followed by the changing of women's role in society. The 19th century witnessed changes in the *perception* of women's nature. Women in this *era* faced difficult and confusing problems from the *fallacy* of what society proposed as their identity and social roles (Koscher, 2006:2). They were restricted to domestic *sphere*. They were expected to *devote* their selves to marriage and motherhood. This situation provided them with only two *options*; living as the society *prearranged* them or choosing to be free of this *oppression* (research artile no. 93).

According to Buku Standar Isi SMP (BSNP, 2006: 124), the *junior* high school students are expected to have linguistic competence. This *linguistic competence* of the students can be seen, among others, through the students' ability in using vocabulary. Brown (2001:315) suggests that vocabulary can be taught within a reading section through class activities that focus the students' attention on vocabulary rather than just learning *isolated vocabulary* by memorizing a list of words. (research article no. 79)

A LFP program developed by Laufer and Nation (1995) which can display the advanced or relatively unusual words in a text analyzed is used to identify the sophistication words. The advanced words are found in BASE LIST 3, UWL, and NOT IN ANY LIST. This program is also integrated with LCA, but the output of LCA only the count results, while LFP displays the advanced words found in the text analyzed.

From the quotations of one paragraph in the undergraduate students no. 93 and no. 79, the advanced words identified are the italic and bold words. From the displays, student no. 93 who gets maximum value employs more advanced words

than student no. 79 (had minimum value). The spread of advanced vocabularies employed by the students above is located in BASE LIST 3, UWL, and NOT IN ANY LIST in the count results of LFP (Laufer and Nation, 1995). In research article no. 93, "*devote, era*, and *perception*" are located in BASE LIST 3; "*sphere*" is found in UWL ; "*fallacy, oppression, prearranged*" are located in ANY LIST. In research article no. 79, "*Junior*" is found in BASE LIST 3, "*isolated*" is found in UWL, while "*linguistic, competence, and vocabulary*" are located in ANY LIST.

Furthermore, description of the employment of lexical sophistication in the undergraduate students' research articles which is described by earlier sophistication measures 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 stuggestions recommended by (Laufer, 1994; Linnarud 1986) that there are different roles of lexical sophistication played in spoken and written proficiency.

Lexical Variation is identified in four different ways, viz., number of different ways, type token ratio, verb diversity, and lexical word diversity. Number of different words are counted using four different measures, namely NDW, NDW-50, NDW-ER50, and NDW-ES50. Type token ratio are computed using six measures, they are TTR, MSTTR, CTTR, RTTR, LogTTR, and UBER.

Verb diversity is measured using 3 measures, they are VV1, SVV1, and CVV1. The last indicator of lexical variation is lexical word diversity. This is calculated using LV, VV2, NV, AdjV, AdvV, and ModV.

Number of different words (NDW) is the first criteria of lexical variation. Essentially the four measures (NDW, NDW-50, NDW-ER50, NDW-ES50) in counting number of different words are developed to get more accurate result. The last three measures are considered transformed measures of NDW. Counting the undergraduate students' 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 resuls shows that the number of different words of the undergraduate students' articles are higher than those of the values of Chinese oral narratives,

The count result produced by one of these measures is provided to show the employment of different words in different research articles. The mean value of number of different words employed in undergraduate students' research articles represented by the result of NDW measure is 677.2. This means the research article contains 677 different word types which are counted at the first time appeared in the research article. The maximum value got by research article no. 93 expresses that this research article employs more different words than research article no. 65 which got the minimum value. Each first paragraph is quoted here from these two research articles to show the differences in the employment of of different words in those paragraphs.

Women's social conditions change from time to time. It is generally followed by the changing of women's role in society. The 19th century witnessed changes in the perception of women's nature. Women in this era faced difficult and confusing problems from the fallacy of what society proposed as their identity and social roles (Koscher, 2006:2). They were restricted to domestic sphere. They

were expected to devote their selves to marriage and motherhood. This situation provided them with only two options; living as the society prearranged them or choosing to be free of this oppression (Arnold, 2005:1) (research article no. 93).

The usage of metaphorical or idiomatic expressions is indispensable if we want to enhance our understanding towards the natural language of English native speakers. Thus, according to Ahmadi and Ketabi (2011), idiomaticity of knguage constitutes a significant role in both laying the foundation and building the comprehension of every language. So, all languages have idioms, that is, a string of words which has distinct meaning from the original one conveyed by the individual words (research article no. 65).

The number of different words counted in research article no. 93 are 62 different words except names of persons. As with the research article no. 65, they are 55 different words. Based on the number of different words employed , a text can be categorized as more complex than other text. The higher the number of different words employed the more complex the lexis is considered.

Type token ratio is the second indicator of lexical variation. In identifying this indicator, six measures (TTR, MSTTR, CTTR, RTTR, LogTTR, UBER) are implemented. Since, the length of research articles is longer than the essay of Chinese learners, the count results compared are taken from the measures that are considered to be able to reduce the sample size effect such as MSTTR and CTTR. The mean value of undergraduate students' research articles using MSTTR and CTTR are 0.72 and 7.76. Compared with the type token ratio of Chinese English learners (0.686 & 4.942), the undergraduate students' type token ratio is higher. This means the undergraduate students employ number of different word types which are higher than those of Chinese learners.

Verb diversity is the third criteria of lexical variation. The measures used to count verb diversity are VV1, SVV1, and CVV1. The last two measures are transformed from the traditional TTR measure. The two measures (SVV1 &

CVV1) are used to describe the employment of verb diversity in the undergraduate students' research articles. The mean values got are 49.73 and 4.92. These values are higher compared with the mean values of Chinese learners (13.415 & 2.556). Description of verb diversity using SVV1 employed by the undergraduate students' research articles can be seen on the paragraphs taken from the research article no. 95 which gets a maximum value (133.12) and from research article no. 60 which gets a minimum value (19.76). Verb diversity in the following quotations is identified through the italic and bold words.

Explicitly, the oath utterance above is **considered** to be assertive of informing. This can be **seen** from the word "Lo!" in which the speaker **gives emphasize** to the truth of proposition of the speaker. This **shows** that the Speaker literally **informs** the hearer about the truth of the dispersion of human effort. However, this is not what actually the speaker **intend** to **convey**. The illocutionary force of the oath utterance is **based** on the context in which S intends to **recommend** H to be generous (in giving charity), dutiful, and **take** in faith of God. From here, it can be shortly **said** that the primary illocutionary act of this oath utterance is directive of **recommending** which is non-literally **expressed** though it has assertive of informing as the secondary illocutionary act (research article no. 95).

The existence of textbooks is an important point and cannot be *separated* from teaching and learning process. Most teachers use textbook in teaching and learning process as a guide for teachers to *explain* the materials inside the textbooks. In teaching and learning process, textbooks have an important role to *provide* a framework of teaching and learning, syllabus, and task. Teachers may *need* more than one textbook in teaching and learning process. Teachers should be selective in *choosing* the appropriate textbooks for their class. To *know* whether the textbooks are appropriate or not, textbook evaluation should be *done* (research article no. 60).

Lexical word diversity is the last criteria of lexical variation which is measured through lexical variation (LV), verb variation (VV2), noun variation (NV), adjective variation (AdjV), adverb variation (AdvV), and modifier variation (ModV). The mean value of the undergraduate students' lexical variation obtained from the ratio of the number of different lexical words to the total number of lexical words in a text is 0.28. Description of the employment of lexical words taken from the first paragraph in research article no. 02 which gets 0.28 is in the following.

Second *language reading comprehension* is *known* as a *highly complex skill* to *master* (Phakiti, 2006). *New vocabularies* and *different text structure* may *cause* the *difficulties* to *foreign language learners* to *master* the language. That is also the *main challenge* for many *students* of ESL. As *stated* by Carlston (2011), one of the *barriers* to *master* reading comprehension is *students' inability* to *engage* the *text* when they do *read*. *Ineffective* reading may *cause* the students to *find* difficulty in *comprehending* the text (research article no. 02).

The italic and bold words in the quotation above express lexical variations employed in the paragraph. In LCA, lexical variation is counted by totalling the number of different lexical words compared to the total number of lexical words.

Verb variation is another indicator of lexical word diversity. The mean value for verb variation is 0.08. Description of verb variation of the research article no. 03 which gets this value is seen in the following quotation.

Writing is an important skill to be *mastered* by language learners because it *reinforces* structures and vocabulary that they have *learned* (Raimes, 1983). In addition, writing *gives* chances to the students to be adventurous with the language and to *go* beyond what they have learned to say. Also, Cohen and Miller (2001) *consider* writing as an important skill since it is an active communicative or social process involving discussion, interaction with the teacher, group work, pair work and peer evaluation (research article no. 03).

The italic and bold words above exhibit the verb variation employed in the paragraph in research article no. 03. LV measures the verb variation by counting the number of lexical verbs to the whole number of lexical words in the research article.

The third measure to count lexical word diversity is NV. The mean value obtained from this measure is 0.25. Description of the employment of noun variation in the first paragraph of research article no. 23 which gets this value is displayed in the following quotation.

The different *pronunciations* or different *sounds* will cause different *meanings*. Speaking without considering the pronunciation will disturb and cause *misunderstanding* in the *meaning* of the *words* spoken. *Phonology* is an *aspect* of *linguistics* which studies the sound *system* of *language* (Crystal, 1985). In phonology, *pronunciation* takes the most important *role*. Pronunciation is the *choice* of sounds used in forming words (Carrel and Tiffany, 1960:4). It is very important to study pronunciation since what we pronounce reflects the meaning of *something*. The different *ways* in pronouncing *phonemes* in English will cause different meanings of *words* (research article no. 23).

The italic and bold words in the quotation above indicate the employment of noun variation in the first paragraph of research article no. 23. Verb variation is calculated based on the number of different lexsical verbs to the total number of lexical words.

The fourth measure to count lexical word diversity is AdjV. The mean value obtained from this measure is 0.06. Description of adjective variation in the research article no. 12 is found in the following quotation.

The italic and bold words above signify the adjective variation employed in the first paragraph in which the research article gets the mean value. The adjective variation is calculated by totalling the number of different lexical adjective to the total number of lexical words in a research article.

The accuracy of a language is determined by the rules of structure as word components, and without the organizational structure imposed on *communicative* purposes, the language would imply be disorganized, as stated by Brown (2007: 420). When we write, we generate our ideas and thoughts then put the ideas into written forms. However, we have to be aware of language grammatical rules so we may convey clearly the messages of our ideas in the written form for a communicative purpose. That is why, it is important for the *foreign* language learners to be taught the structures or grammar especially in composing their own writing production. The considerations of learning grammar by DEE and QCA (1991: 21, in Eve English and John Williamson, 2005: 69) lie on how word choices and orders are *crucial* to meaning, the nature and use of nouns, verbs, and pronouns, how ideas may be linked in sentences and how sequences of sentences fit together. Haudeck (1996, as cited in Al-Mekhlafi & Nagaratnam, 2011:72) has reported that many learners have difficulty in internalizing grammar rules, although these have been taught intensively. Lado"s (1957:59, as cited in Bennu, 2012:1) theory indicates that those structures that are *different* will be *difficult* because when transferred they will not function satisfactorily in the foreign language and will therefore have to be changed (research article no. 12).

The fifth measure to count lexical word diversity is AvdV. The mean value obtained from this measure is 0.02. The description of the implementation of adverb variation of a paragraph in research article no. 34 represents the research article which gets the mean shown in the following quotation.

With regard to the types of texts taught in junior high schools in Indonesia, past tense is an important part of language features of, *particularly*, recount texts taught to the 8th and 9th graders. However, students still find difficulties in using past tense. The researcher assumes that students in Indonesia *commonly* make errors in using past tense because there are some differences between *Bahasa Indonesia* and English when we talk about something in the past. In English, if we want to talk about something in the past, we do not only put the adverb of time, but also change the verbs into the past form. Therefore, we will know whether someone is talking about the past, present or future by knowing the verb he/she uses without looking at the adverbs of time. On the other hand, in *Bahasa Indonesia*, the only way for us to know whether someone is talking about the past, present or future is by knowing the adverb of time (research article no. 34).

The adverbs employed in the paragrph quoted above are *commonly* and *particularly*. Adverb variation is counted by totalling the number of different adverb to the total number of lexical adverb in a research article.

The sixth measure to count lexical word diversity is ModV. The mean value obtained from this measure is 0.08. This value expresses the total number of adjective and adverb as modifiers compared with the total number of lexical words in the students' research articles. As a description on the employment of modifier in the students' artiles, one paragraph is quoted from the research article no. 40 which represents the research article which gets mean value as shown in the following.

For *beginner* learners, *especially* Junior High School students, teacher should present new vocabularies through some *fun* activities that can help the students memorize and retain the vocabulary that has been taught. Because young learners' *attention* span is *naturally* short, they will pay attention to something of their interest. Thus, efforts must be made to arouse and maintain their interest, for example by involving them *physically*, *verbally*, and *mentally* by using games, *role* plays, songs, and simulations. In addition to this, the use of various types of media, such as visual, *audio*-visual, audio are *highly* desirable for them. Based on
the reasons, teachers should use *appropriate* method and media to increase the students' motivation, interest and achievement in learning vocabulary (research article no. 40).

The italic and bold words above expressed the employment of adjective and adverb as modifiers. In this case, the adjectives modify noun, and the adverbs modify adjectives and verbs. For example: fun activities, naturally short, and involving them physically, etc.

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.

5.2 The Undergraduate Students' Syntactic Complexity

The first discussion deals with the undergraduate students' mean length of production unit. As found in chapter IV, the mean length of production unit is counted using three measures, i.e. MLS, MLT, and MLC. 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 WECCL (Written English Corpus of Chinese Learners) and of NSs' argumentative essays in LOCNESS (Louvain Corpus of Native English Essays).

The length of prodution 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).

The mean length of sentence is identified by the number of words in a sentence. In other words, the research article with maximum value (28.716) containes sentence which consist of at least 28 words, while the research article with a minimum value (14.626) contains at least 14 words. To see the description of the sentences employed by the undergraduate students, two sentences from each article with maximum and minimum values are described below.

Considering the difficulty of mastering reading skill, there are many techniques applied by teachers to encourage the students to be able to make sense of the written words (research article no. 68).

Richards and Renandya (2002:273) affirm that reading receives a special focus since the written texts serve various pedagogical purposes in which comprehending the written texts can enhance the process of language acquisition (research article no. 68).

Nowadays, many students have learned English in their school since English is a global language (research article no. 69).

In the eight grade, two types of texts taught to the students are recount and narrative texts (research article no. 69).

The first and the second quotations above are the examples of a sentence which consists of 28 words, while the third and fourth quotations comprise of a sentence which contains 14 words.

The mean length of T-unit is identified by the number of words in T-unit.

A T-unit is one main clause plus any subordinate clause or non clausal structure

that is attached to or embedded in it (Lu, 2010). Description on the length of Tunit in the students' research articles is represented by the research articles with maximum (26.808) and minimum (14.047) values.

Students need English teachers who have outstanding skills, abilities, and competences in both their knowledge of English as well as in their professional life (research article no. 70).

This activity was very important for the English teachers since they should maintain their teaching materials 'fresh' so that students could understand the lessons better (research article no. 70).

The quotations above are taken from the research article which has maximum value as the longest T-unit (26.808). The first T-unit counted by Stanford Parser consists of 27 tokens or 27 words. The T-unit consists of one main clause and one subordinate clause. Based on the count of Stanford Parser, the second T-unit contains 27 tokens or 27 words and comprises of one main clause and two subordinate clauses.

Meanwhile there is one student from the regular class who receives the same score (research article no. 69).

The reading strategies applied by the regular teacher are the modeling and grouping strategies (research article no. 69).

The next two quotations above are taken from the research article which obtains minimum value as the shortest T-unit (14.047). The first T-unt based on the count of Stanford Parser embodies 14 words in one main clause and one subordinate clause. The second T-unit based on the count of Stanford Parser also comprises of 14 words. This T-unit consisted of one main clause and one subordinate clause.

Mean length of clause is described by the number of words in a clause. Clause is characterized by a structure with a subject and finite verb and includes independent clauses, adjective clauses, adverbial clauses, nominal clauses (Lu, 2010: 481). The maximum value (14.525) is got by research article no. 55 and the minimum value (8.245) is obtained by research article no. 64.

Furthermore, it was also expected that the students could use the materials independently with the minimum guidance of the teachers because *the answer keys and audio recording were created as a means of self-study materials* (research article 55).

Hirsi Ali also went to Quran school where *she had the chance to discuss religion a lot* (research article no. 64).

The quotation above is a sentence which comprises of one main clause and two subordinate clauses. The italic and bold words are a clause which its length consists of 14 words. In the second quotation, a sentence taken is a main clause plus subordinate clause which consists of 9 words since a clause contains eight words is not found in the research article.

The second point of discussion pertains to sentence complexity which is indicated by C/S. 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: 490). This shows that the sentence complexity of the undergraduate students' research articles is higher than that of Chinese learners' argumentative essays.

Sentence complexity is identified through the number of clauses per sentence. Description of the research articles which gets maximum value (3.209) and minimum value (1.439) on sentence complexity are represented by research articles no. 138 and 60 as follows. The former quotation consists of three clauses and the latter comprises of one clause.

The double life here means that he has to go on with the marriage plan to May, while he keeps imagining to be with Ellen (research article no. 138)

Teachers may need more than one textbook in teaching and learning process. (research article no. 60)

The third discussion goes to the Amount of Subordination which are indicted by C/T, CT/T, DC/C, and DC/T. The mean values of the undergraduate students' research articles using C/T and CT/T are 1.752 and 0.470. 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: 490). Meanwhile, the mean values of the undergraduate students' research articles produced by DC/C and DC/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: 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.

Complex T-unit in the following quotation is taken from the research article no.1, which gets mean value 0.470 or 0.473. Complex T-unit is a T-unit that contains a dependent clause (Lu, 2010). The dependent clause in the complex T-unit below is the italic and bold words

Establishing English extracurricular activities is considered as advantageous for the students *as they have more time and exposures to learn English* (research article no. 1).

The fourth discussion relates to the amount of coordination. The mean values of the undergraduate students' amount of coordination using CP/C, CP/T,

and T/S are 0.299, 0.561, and 1.007. On the other hand, the mean values of NNS-High amount of coordination are 0.231, 0.365, and 1.126 (Ai & Lu, 2013: 258). While the mean values of NS' amount of coordination are 0.254, 0.430, and 1.121(Ai & Lu, 2013: 258). Comparing the mean values of undergraduate students' research articles, NSS-High, and NS, the undergraduate students' research articles have higher amount of coordinate phrase in clause and in T-unit than NNS-High and NS. However, based on the amount of T-unit per sentence, NNS-High is the highest, then NS and the udergraduate students' research articles.

The maximum value got form CP/C is 0.681. This value is represented by research article no. 10. Coordinate phrase in clause "... who have different interests and needs", in the following quotation is the words itilicized and bolded.

The maximum value got from CP/T is 1.063. This value is also represented by no. 10. The following quotation is a T-unit that consists of a coordinate phrase that is italicized and bolded.

For the students, a textbook affects *their motivation and performance* through the lessons (research article no. 10).

The maximum value got from T/S is 1.270. This value is represented by research article no. 68. A sentence quoted from research article no. 68 describes a complex sentence which consists of one T-unit, shown by italic and bold words below.

In addition, Cunningsworth (1995:136) states that there is no textbook that can cover all of the materials for students who have *different interests and needs* (research article no. 10).

Even though reading is learnt from an early stage, *it is often found difficult for many readers*. (research article no. 68).

The fifth discussion refers to degree of phrasal sophistication which are counted using CN/C, CN/T, and VP/T-unit. The mean values of udergraduate students' degree of phrasal sophistication were 1.397, 2.441, and 2.392. Using CN/C and CN/T, the undergraduate students' degree of complex nominals per clause, complex nominals per TH-unit are the highest followed by NS (1.222 and 2.089) and NNS-High (1.064 & 1.669) (Ai & Lu, 2013: 258). Using VP/T, the undergraduate students' degree of verb phrase per T-unit is higher than Chinese Learners level III (Lu, 2010: 490).

Complex nominals consist of (i) noun plus adjective, possessive, preposition phrase, relative clause, participle, or appositive, (ii) nominal clauses, and (iii) gerunds and infinitives in subject position (Cooper, 1976 in Lu, 2010). Degree of phrasal sophistication is observed through the presence of complex nominals per clause and per T-unit and verb phrase per T-unit. For the easiness of identification, the implementation of coordinate phrase in clause and T-unit as well as verb phrase in T-unit, the examples are taken from the articles which got maximum values.

The maximum value got from CN/C is 2.000 and from CN/T is 4.000. These values are obtained by research article no. 96. The first complex nominal italicized and bolded is found in T-unit *"the key stone is created by the Priory in the past couple of decade"*. The second complex nominal is found in clause *"he does not mention the exat location"*. Both clauses in the following quotation are complex nominals which consist of noun plus adjective. In the dialog above, Langdon states that the keystone is created by the Priory in the *past couple* of decades but he does not mention the *exact location* of it (research article no. 69).

The maximum value got from VP/T is 3.836. This value is obtained by research article no. 138. The following complex sentence contains two T-units. The first verb phrase is found in T-unit "*The existentialist concept of self-determination should be put above the super egoic*" and the second verb phrase was found in T-unit "*The existentialist concept of self-determination has been underlying* in the conclusions of the studies done upon Wharton's The Age of Innocence"

The existentialist concept of self-determination which *should be put* above the superegoic *has been underlying* in the conclusions of the studies done upon Wharton's The Age of Innocence (research article no. 38).

The above explanation reveals that the length of production unit of the undergraduate students (MLS, MLC, MLT) are higher than those of NS. This is different with what is found by Hinkel (2011:529) who reveales that NNS relies on shorter sentences and clauses (T-unit) with fewer words per clause. The same occurrence happens to the undergraduate students' degree of phrasal sophistication which is higher than those of NS.

The amount of subordination of the undergraduate students is lower than those of NS. This is similar with what is found by Hinkel (2005: 621; 2011) and Sylva (1993) that NNS tend to use less subordination than NS.

On the other hand, the amount of coordination of the undergraduate students are higher than those of NS. This is also in line with what Hinkel (2005: 621) and Sylva (1993) find that NNS tends to use more coordination than NS.

5.3 The Quality of the Undergraduate Students' Research Articles

As explained in chapter IV, the mean value of the quality of the undergraduate students' research articles is 3.5. The number of research articles who gets 3.5 are 24 research articles, namely, 2 students (No. 3 & 21) obtain 2 for flawless language, but obtain 4 for convincing rethoric, retrieving academic insight, and elegant style or in short pattern (2, 4, 4, 4). 8 students (No. 4, 24, 38, 39, 45, 59, 122, 128) obtain (3, 4, 4, 3); 1 student (No. 5) obtains (4, 3.5, 3.5, 3); 1 student (No. 6) got (3, 4, 3.5, 3.5); 1 student (No. 17) obtains (4, 3, 3, 4); 1 student (No. 20) obtains (4, 3.5, 3, 3.5); 4 students (No. 41, 54, 86, 137) obtain (4, 3, 4, 3); 4 students (No. 48, 100, 130, 131) obtain (4, 4, 3, 3); 1 student (No. 104) obtains (3, 3, 4, 4); 1 student (No. 106) obtains (3, 4, 3, 4).

Value 2 received by the student no. 3 in *flawless language* is caused by some words are written with other words continuously in which should be written separately.

The student's Writing proficiency who obtains 3 in *convincing rethoric* is represented by student no. 17. The student's research article is entitled "The Teaching of ESP 'Mojokerto Toursim and Industry'class at SMK PURI Mojokerto" The analysis on this research article is as follows:

In the introduction, the first point to explain is *establishing a research* territory. In her introduction the student explains the importance of the topic that the teaching of English now is directed to the means of language, that is to teach the language for the purpose of using the language to communicate. The next point explained is the content of ESP teaching which covers topic, material, method, and evaluation. This explains the concept of teaching ESP according to

Hutchinsong. This explanation is examined as establishing the research territory by reviewing concept as it is explained by Basthomi (2009) that establishing a research territory is still mostly done by the Indonesian writers. Establishing a research territory is mainly done by reviewing on the previous research of related topic even though reviewing concept does not mean it is not important (Basthomi, 2009).

In *Establishing a niche*, the student does not explain the gap between her search and the previous research because of not doing a review which contribute in determining the reason for the topic investigated. In *Occupying the niche*, the student does not explain explicitly the nature of the research she does, and so does the structure of her writing.

In writing the *research method*, the student begins her explanation by mentioning the design of her research and the instruments she uses to gather data without explaining the aims of those instruments. The student explaines that the sampling is done randomly, by mentioning the population and the samples but not explaining the criteria in taking the samples. Further, the student does not explain explicitly the model of analysis she implements.

In the result, the student gives a complete description on the syllabus used, the material employed, the methods implemented, and the assessment done. However the source referred to the data analysis are not mentioned, and less evaluation given on the description mentioned.

In the Discussion, the student has appropriately explained the reason for the decision of determining the syllabus in the school observed. She also has mentioned the reason for various methods that are employed, but more additional

analysis on the methods used by the teachers is still needed. Further, the student also needs to give more explanation about the materials that have been developed by the teachers and to the extent the materials have been in line with the theory revealed. Lastly, a simple explanation on the assessment done by the teachers still needs a concrete examples especially on the match of the assessment with the other three aspects explained before. While in the *Conclusion*, the student has answered the first, second, third, fourth research questions well.

Value 4 in *Retrieving Academic Insight* is described by student no. 104 whose research article is entitled "The Use of English in Teaching Mathematics and Science at R-SMA-BI 1 Situbondo". In assessing this research article based on the criteria *Retrieving academic insight*, the six components of a research article are involved since these components are integrated each other. In the following, it is described the quality of the student' research article which obtained 4.

In the *abstract*, the student mentiones the aim the study but does not explain the procedure of doing the research. The findings have been stated explcitely.

In the *introduction*, the rational for the teacher of Mathematics and Sciences using English to teach are depicted well, but the problem of research and the ways to solve are not available as well as the announcement of the research purpose. *In the research method,* the reason for choosing the source of data and the reason for choosing the subjects are stated logically. The instruments and the aim of instruments are explicitely mentioned. The ways of analyzing the data are

displayed orderly.

In the *findings*, the answers to the research questions are clearly stated and explained in order.

In the *discussion*, the findings in the previous chapter are explained and correlated with the previous research.

In the *conclusions and recommendation, conclusion consists* of summary of results and discussion. The recommendations have been related to the findings and discussion which give solution to the problems found in the research.

Value 4 in *Elegant Style* is obtained by the student no. 106. The title of her journal is "Developing a Board Game for Speaking Activities of Grade VIII Juniour High School Students".

The criteria used for elegant style is the intensive use of lexical and syntactic complexities which support comprehensible meaning. Based on the count result shown by LCA, the lexical density employed by the student no. 106 is 0.52. This value is categorized as average value. The lexical sophistication of student no. 106 using LS1, LS2, VS1, VS2, CVS1are 0.18, 0.29, 0,05, 1,51, and 0.87. Compared with other values of research articles, the value got by student no. 106 from LS1 is categorized as minimum value. The value produced by LS2, VS1, VS2, and CVS1 are categorized as below mean value. Student no. 106 of lexical variation based on number of different words using NDW, NDW-50, NDW-ER50, and NDW-ES50 are 640, 42, 37.20, and 35.70. Compared with other research articles, the value given by NDW, NDW-ES50 are categorized as nearly average, while the value given by NDW-50 is higher than the mean value.

Lexical variation based on type token ratio, the value obtained by Student no. 106 using TTR, MSTTR, CTTR, RTTR, LogTTR, Uber are 0.17, 0.70, 7.27, 10.29, 0.78, and 16.47. Compared with other values of the research articles, all the values got from these measures were categorized below average. Other lexical variation based on verb diversity, the value obtained by Student no. 106 using VV1, SVV1, and CVV1 were 0.28, 44.33, and 4.71. Compared with other values of the research articles, the values got by these measures categorized student no. 106 as below average. The last criteria are using LV, VV2, NV, Adj.V, Adv.V, and Mod.V.The values got by research article no. 09 are 0.26, 0.08, 0.22, 0.05, 0.02, and 0.07. Using VV1 and Adv.V, research article no. 106 gets average value, whilst the rests go below average.

In relation to the syntactic complexity of the research article of Student no. 106, the mean length of production unit of her research article using MLS, MLT, and MLC are 21.455, 20.423, and 11.006. These values are above average for MLS and MLT, and below average for MLC. The sentence complexity of the research article of the student no. 109 was 1.949. The complexity of this value is above average. The amount of subordination of the research article no. 106 using C/T, CT/T, DC/C, and DC/T are 1.856, 0.508, 0,406, and 0,754. All these values are categorized as above average compared with other research articles. The amount of coordination of the research article no. 106 using CP/C, CP/T, and T/S are 0.297, 0.551, and 1.051. These values are categorized as below average for CP/C, and above average for CP/T and T/S. The values of degree of phrasal sophistication for the research article no. 106 using CN/C, CN/T, and VP/T are

1.144, 2.123, and 2.840. These values are categorized below average for CP/C and CP/T and above average for T/S.

Based on the count result of Lexical Complexity Analyzer, research article no. 106 gets average on lexical density, minimum and below average on lexical sophistication, nearly average and above average on lexical variation based on number of different words, below average on lexical variation based on type token ratio and verb diversity, average and below average on lexical variation based on lexical word diversity.

Based on the count results of Syntactic Complexity analyzer, research article no. 106 obtaines above and below average based on mean length of production unit, obtaines above average based on the amount of subordination, obtaines above average based on sentence complexity, obtaines below and above average based on the amount of coordination, and obtaines below and above average based on degree of phrasal sophistication.

Relating to the calculations above, research article no. 106 should just get 2 for the criteria of elegnt style.

5.4 Correlation between the Undergradulate Students' Lexical and Syntactic Complexities and the Quality of their Writing

Lexical complexity which covers three multi dimentional 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).

Lexical density which is defined as lexical words or content words such as noun, verb, adjective, and adverb loads important information required for understanding. Noun tells us which object, verb informs the action happening, adjective gives a detail to object or person, and adverb explains how, when, and where the action happens. The wide use of lexical words in a text provides for the potentiality of message or ideas to declare.

This current research finds that there is no correlation between lexical density and the quality of students' research articles. In line with this finding, Engber, 1995 and Linnarud, 1986 in Lu, 2012 in their L2 writing studies have reported that lexical density appeares not to relate with the quality of writing. The description on the no correlation between both, research article no. 4 is taken as an example. This research article gets mean value on its quality, that is 3.5. However, the result got from LD which describes the mean value of the students' lexical density is 0.52. The value is categorized as average. On the other hand, the research article which obtaines maximum value (research article no. 110) just gets 3.250 which is under the mean value of the research articles. So there is a mismatch between the lexical density gets by the students and the quality of their research article. However, there is also an in line values of lexical density and quality of research article, such as research article no. 121 which got maximum value on lexical density, and got 4 on quality of research article.

Lexical sophistication is defined as advanced words or relatively unusual words. For advanced students, using LFP, these words are found in BASE LIST 2,

BASE LIST 3, UWL, and NOT IN ANY LIST (Laufer & Nation, 1995). List 2 and 3 consists of a wide range of content words above 1000 frequent words. UWL is a list of 836 word families containing vocabularies which are not found in Base List 2 but are frequently and widely found across a variety of written academic texts from a variety of disciplines. NOT IN ANY LIST is group of words which are not found in list 2, 3 and UWL. The intensive appearance of these words in students' research articles categorizes the research articles equal with advanced academic texts.

In relation to the appearance of these words with the quality of research articles, there is no correlation found. This result differs from ones reported in previous L2 writing studies which report that lexical shopistication characterizes L2 written texts (Laufer, 1994 & Linnarud, 1986). Compared with the mean value of the research article quality (3.5) which is categorized as just below the criteria of great article, the students' mean of lexical sophistication is higher than that of Chinese learners. In other words, the undergraduate students' research article containes more sophisticated lexis than those of Chinese learners' oral narratives.

Lexical variation is commonly identified by the variety of different words employed in a text. Lexical variation employed describes the writers' lexical proficiency. A wide range of lexical variation provides a writer with a help to explain ideas or message properly.

In line with the statement above, this present research shows the existence of correlation between the wide varieties of different words and the quality of research articles. The count result of NDW and the quality of research articles have low and positive correlation. The result shows that the research article

which get the highest value on the number of different words gets 3.75 on quality of research article. This value describes the value which is above the average value of quality of the whole articles.

Different with NDW as a measure of lexical variation, TTR as another measure of lexical variation finds that the count result of TTR has negative and significant correlation with the quality of research articles. As a comparison for this correlation, research article no. 17 which gets mean value on quality, gets 0.23 value from TTR measure. The value from the TTR is categrized as above average or the value explains that 23 percents of the number of tokens in research article no. 17 are different word types. This value is lower than that of Chinese learners' spoken narratives. Based on this comparison, the number of type token ratio according to the count result of Pearson product moment correlation should be decreased in order to achieve higher quality seems not irrelevant with the value of number of type token ratio which is still low compared with Chinese learners' spoken narratives.

Verb diversity is the third criteria of lexical variation which is counted based on the number of different word types to the total number of word types in a research article. The count result of verb diversity measure has negative and significant correlation with the quality of research articles. This means the improvement of quality of research articles is followed by decrease in number of verb diversity. This result is not in line with previous research that lexical variation increases in number on higher level of written texts (Laufer & Nation, 1995) and lexical variation is always found in students' written academic text (Lu, 2012). Furthermore, this value is lower than that of Chinese learners' spoken

narratives. It means the verb diversity value is still low, so it needs to increase, not to decrease in order to achieve higher quality.

Lexical word diversity is the fourth criterion of lexical variation which is identified based on lexical word variation, verb variation, noun variation, adjective variation, adverb variation, and modifier variation. Except adverb variation, lexical word, verb, noun, adjective, and modifier variations have negative and significant correlation with the quality of research articles. Compared with Chinese learners' spoken narratives, the values of these five indicators are lower than that of Chinese learners' spoken narratives. The findings find that since the lexical word, noun, verb, adjective, and modifier variations are still low, number of these variations are not logically decreased in order to achieve higher quality as the coeficient correlation told. The only lexical variation value which is higher than that of Chinese learners' spoken narratives is the value of adverb variation but the correlation is not significant.

Syntactic complexity is identified based on mean length of production unit, sentence complexity, amount of subordination, amount of coordination, and degree of phrasal sophistication (Lu, 2010).

The mean length of production unit of the undergraduate students' articles covering: length of sentence, length of clause, and length of T-unit. The length of these production units is not significantly correlated with the quality of research articles. As a descritption, 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 obtaines 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. However the mean length of the students' production unit is longer than the length of production unit of NNS-High in WECCL 2.0 (Written English Corpus of Chinese Learners) and even longer than NS (essays written by American University students in LOCNESS). These results are different with what is found by Hinkel (2003, 2005, 2011) who reveals that NNS relies on shorter sentences and clauses (T-unit) with fewer word per clause.

Some assumptionts appeared on the result of comparison between the length of prduction unit of the undergraduate students and the American university students. The assumptions were referred to firstly, the undergraduate students' (NNS) tendency to use more words to explain their ideas (Hinkel, 2011) and secondly, the different level of complexity between the corpus of Chinese learners which were in the form of essays as concordances in the software and the undergraduate students' research articles which were in the form of research articles.

Sentence complexity which is characterized by number of clauses per sentence has no correlation with the quality of research articles. The highest score (4) gets 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).

Compared with the mean values got in writing argumentative essays by NNS-High in WECCL corpus and by American university students in LOCNESS, the amount of subordinations of the undergraduate students' research articles are higher than NNS-High but is lower than NS. This result is in line with what is found by Hinkel (2003, 2005, 2011) that NNSs tends to use less subordination in their written texts.

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.

The opposite of amount of subordinations, according to Hinkel, NNS tends to employ more coordinations which reversed to NS's who tends to use less

coordination in their written texts. In this study, the mean value of coordinate phrases in clause and in T-unit are higher than those of NNS and NS. However, the undergraduate students' amount of T-units in sentence are lower than NNS's and NS's.

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 obtaines 3.00 on quality.

Compared with the mean value of NNSs'-High and NSs', degree of complex nominals per clause (CN/C) and per T-unit (CN/T) of undergraduate students is higher than NNS's and NS's and degree of verb phrases per T-unit is higher than NNS's.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

This chapter presents conclusions of the findings and discussion and reveals some recommendations for some parties relevant with the findings of this study.

6.1 Conclusions

This study investigates the trends of lexical and syntactic complexities in the undergraduate students' research articles and their correlation to their quality. To decide whether the scores of lexical and syntactic complexities were high or not, comparative data were needed since the output of the softwares were only the count results in the forms of numbers without displaying the lexical and syntactical items of the complexities. Similar comparisons on lexical sophistication between spoken and written proficiency had ever been done by Laufer, 1994 and Linnarud, 1986. The lexical complexity of the undergraduate students' research articles was compared to the lexical complexity of Chinese learners' oral narratives, while the syntactic complexity was compared to the syntactic complexity of argumentative writing of NNs-High and NS in WECCL and in LOCNESS.

Lexical complexity of the undergraduate students articles was identified through its lexical density, lexical sophistication, and lexical variation. Lexical density and lexical sophistication of the undergraduate students' research articles were higher than those of Chinese learners but did not go along with the quality of research articles yet. Lexical variation of the undergraduate students' research

articles based on the number of different words employed was higher than that of Chinese learners' oral narratives and had a positive and low correlation with the quality of research articles. The number of type token ratio, using MSTTR and CTTR and verb diversity using SVV1 and CVV1, of the undergraduate students' research articles were higher than those of Chinese learners' oral narratives.

Lexical word diversities including lexical variation, verb variation, noun variation, adjective variation, and modifier variation were lower than that of Chinese learners' oral narratives. Even though the negative correlations of these indicators with the quality of research articles were significant, it was not logical for the undergraduate students to decrease the employment of these indicators to achieve higher quality on the research articles, since the values of the indicators themselves were lower than those in Chinese learners' spoken narratives. Furthermore, adverb variations of the undergraduate students in Chinese learners' oral narratives and had no correlation with the research article quality.

Syntactic complexity was characterized by mean length of production unit, sentence complexity, amount of subordination, amount of coordination, and degree of phrasal sophistication. Mean length of production unit had no correlation with the quality of research articles however, the length of production unit was higher than NNS's and NS's. 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 sentence complexity of the undergraduate students was not in line with the value of research article quality. However, the value of sentence complexity was higher than that of Chinese learners' essays.

The amount of subordination had no correlation with the research article quality, however, it was higher than argumentative essays of NNS-High but lower than those of NS. This finding was similar to what Hinkel (2003, 2005, 2011) found that NNS used less subordination.

The amount of coordination of the undergraduate students' research articles which were determined by the mean values of coordinate phrase in clause and in T-unit was higher than those of NNS and NS and by the value of the amount of T-units in sentence which was lower those of NNS and NS. The former values were in line with Hinkel (2003, 2005, 2011) who stated that NNS tended to use more coordination than NS, while the value for the amount of T-unit was not.

Degree of phrasal sophistication of the undergraduate students was shown by the degree of complex nominals per clause and per T-unit which was higher than NNSs' and NSs' and degree of verb phrases per T-unit which was higher than NNS's. The values were not in line with the quality of research articles, or in other words, the high quality of the undergraduate students was not followed by high value of phrasal sophistication degree or vice verse.

The quaility of udergraduate students' research articles got 3.5 which was categorized as above good. This indicated that the quality of the undergraduate students' articles nearly achieved or only 0.5 point needed to achieve the criteria of great articles. Based on the employment of lexical complexity in the research articles, this research found that number of different word had a positive and

significant correlation but low to the quality of research articles. Whereas on the employment of lexical word diversity, the undergraduate students need to improve the employment of lexical word variation, noun variation, verb variation, adjective variation, adverb variation, modifier variation. Related to the employment of syntactic complexity in research articles, in order to resemble the employment of syntactic complexity with NS, the undergraduate students need to improve the employment of subordination (dependent clause per clause and per T-unit) and to decrease the employment of coordination (coordinate phrases in clause and in T-unit).

The no correlation between lexical and syntactic complexities in undergraduate students' research articles and their quality were shown by the no consistency on some research articles on the employment of lexical and syntactic complexities to the quality of research articles. It was seen from the highest values got by some research articles but they were not followed by highest values on lexical and syntactic complexities and vice verse.

6.2 Recommendations

Having reviewed the results of data interpretation and compared with the result of previous research, some recommendation are proposed in the following.

For the lecturers of Complex English Grammar and Academic Writing, the result of this research informs the lexical and syntactic complexities in the undergraduate students' research articles which are needed to be improved. 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, 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 suborodination, and less coordination in their academic writing as well as lexical and syntactic complexities in general.

For the undergraduate students, lexical and syntactic complexities characterized the elegant style of advanced academic texts so that it is recommended to the students to be aware with them especially those that their employment still should be improved in order the undergraduate students' style in writing academic texts achieves the elegant style of national and international journals

For future researches, the trends of undergraduate students' lexical and syntactic complexity would be more real if it is compared with other national and international journal research articles. Through this comparison, it can be mapped the trends of lexical and syntactic complexities use of the undergraduate students' research articles that should be improved and retained compared with those journals.

For the software developers, it is mostly accepted that the presence of softwares are very important in doing research and in teaching so that in order the softwares can be widely used, it is very important to consider the easiness of access and operationalization of softwares.

REFERENCES

- Afini & Cahyono, B.Y. 2012. The Lexical Richness in Narrative Texts of XI Graders of MAN 3 Malang. *English Language Education*, 1(1). (Online), (http://journal-online.um.ac.id), accessed February 23, 2013.
- 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 Concortium (CALICO-10). (Online), (http://www.personal.psu.edu/hua126/papers/calico10_ai_lu.pdf), accessed February 23, 2013.
- 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.
- Anwar, K. 2010. *Rethorical Pattern of Research Articles in Languge Teaching Journal*. Unpublished Dissertation. Malang: PPS UM.
- Akuntono, I. 28 Februari 2012. Kewajiban Publikasi Ilmiah Dibatalkan? *Kompas.com.* (Online), (Http://edukasi.kompas.com/read/2012/02/28/08094149), accessed Feb. 17, 2013
- Apriani, R. 2011. Propositions on Reading Texts of upper Secondory English Textbooks: a Corpus-based Study. Unpublished Thesis. Malang: PPs UM.
- Arya, D.J., Hiebert, E.H. & Pearson P.D. 2011. The Effect of Syntactic and Lexical Complexity on the Comprehension of Elementary Science Texts. *International Electronic Journal of Elementary Education*, 4(1), 107-125, (Online),

(www.academia.edu/attachments/33748708/download), accessed March 20, 2013.

- Baker, P. 2011. Corpus Method in Linguistics. In Litosselit, L. (2nd ed.). *Research Methods in Linguistics*. London: Continuum International Publishing Group.
- Barrrot, J.S. 2012. Exploring the Lexical and Syntactic Features of the Learners' Narratives. *The Philippine ESL Journal*, 9: 38-55, (Online), (http://www.philippine-esl-journal.com/V9-A3), accessed September 05, 2013.
- Basthomi, Y. 2006. The Rethoric of Research Article Introductions Written in English by Indonesians. Unpublished Dissertation. Malang: PPS UM.
- Basthomi, Y. 2009. Examining Research Spaces in Doctoral Prospectus. *TEFLIN Journal*, (Online), 20 (2): 140-157, (http://journal-online.um.ac.id), accessed September 05, 2013.
- Brown, H.D. 2007(3rd ed.). *Teaching by Principles: an Interactive Approach to Language Pedagogy*. New York: San Fransisco State University.
- Bromwich, W. 2008. Blood, Toil, Tears and Sweat? Achieving Proficiency in Academic Writing. Working Paper *ADAPT*. (Online) (http://moodle.adaptland.it/pluginfile.php/4930/mod_resource/content/1/ 329WP_08_71.pdf), accessed March 23, 2013.

- Caspi, T. 2010. A Dynamic Perspective on Second Language Development.
 Groningen Dissertation in Linguistics 85. ISSN 0928-0030. ISBN 978-90-367-4526-0. Groningen: University of Groningen. (Online)
 (http://irs.ub.rug.nl/ppn/329338412), accessed March 23, 2013.
- Ellis, R. 2001. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Farahani, A.A.K. & Meraji, S.R. 2011. Cognitive Task Complexity and L2 Narrative Writing Performance. Journal of Language Teaching and Research, 2 (2): 445-456, (Online) (http://ojs.ac.ademypublisher.com/)
- 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*, (Online) (academics.utep.edu/Portals/321/.../Francis/FrancisRomoGelman02.pdf), accessed March 01, 2013.
- Haan, P.D. & Esch, K.V. 2008. Measuring and Assessing the Development of Foreign Language Writing competence. *PORTA LINGUARUM* 9: 7-21, (Online), (http://digibug.ugr.es/bitstream/10481/31742/1/Haan.pdf), accessed March 01, 2013.
- Hesamouddin, S.A., Behzad, G., & Azar, H.F. 2012. Analyzing Research Article Introductions by Iranian and Native English Speaking Authors of Applied Linguistics. *International Journal of Research Studies in Language Learning*, 2 (3): 3-18 (Online), (profdoc.um.ac.ir/articles/a/1030442.pdf), accessed March 01, 2013.
- Hinkel, E. (2003). Simplicity without Elegance: Features of Sentences in L2 and L1 Academic Texts. *TESOL Quarterly*, 37: 275–301. (Online), (http://www.elihinkel.org/dowloads.htm), accessed March 01, 2013.
- 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. (Online) (http://www.elihinkel.org/dowloads.htm), accessed March 01, 2013.
- 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: Routlege. (http://www.elihinkel.org/dowloads.htm), accessed March 01, 2013.
- Kitamura, M. 2012. Development of Japanese EFL College Students' English Writing Skills: An Examination of Subordinators in Essays. (Online), (cblle.tufs.ac.jp/assets/files/publications/working_papers.../325-336.pdf), accessed February, 28 2013.
- 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. @Oxford University Press.

(Online), (*www.pebc.org/.../Larsen-Freeman-Complxty-Fluncy-Accrcy-in-Oral-Wt...*)

- Latief, M.A. 2010. *Tanya Jawab Metode Penelitian Pembelajaran Bahasa*. Malang: UM Press.
- Latief, M.A. 2012. *Research Methods on Language Learning (An Introduction)*. Malang: UM Press.
- 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. Oxford University Press.
- 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 ISSN 1384–6655 / E-ISSN 1569–9811 © John Benjamins Publishing Company
- 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, accessed November 14, 2013.
- Lu, X. 2012. The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*. 96(2): 190-208.
- Mukminatin, N. 1997. *The Differences of Students' Writing Achievements across Different Course Level.* Unpublished Dissertation. Malang: PPs UM.
- 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. (Online), (www.ub.edu/GRAL/pubs/Celaya&Naves2009Writing.pdf).
- Nuh, M. 13 February 2012. Mendikbud: Publikasi di Jurnal Ilmiah Jalan Terus. *Kompas.com*. (Online),

(http://entertainment.kompas.com/read/2012/02/10/2212575)

- Pedoman Penulisan Karya Ilmiah. 2010 (5th ed.). Malang: Universitas Negeri Malang.
- Pritomo. 2012. Lexical Richness in Teacher Talk of a Non Native English Teacher in the Foreign Language Classroom. Unpublished Thesis. Malang: PPs UM.
- Safnil. 2013. A Genre-based Analysis on the Introductions of Research Articles Written by Indonesian Academics. *TEFLIN Journal*, (Online) 24 (2): 180-200, (http://journal-online.um.ac.id), accessed January 28, 2014.
- Šišková, Z. 2012. Lexical Richness in EFL Students' Narratives. University of Reading: Language Studies Working Papers. ISSN 2040-3461, 4: 26-36. (Online) (http://www.readingconnect.net/web/files/english-language-andliterature/elal_lswp_vol_4_siskova.pdf), accessed 13/04/2013.
- Stockwell, G. 2005. Syntactical and Lexical Development in NNS-NNS Asynchronous CMC. *The JALT CALL Journal*, 1 (3): 33-49. (Online)

(http://journal.jaltcall.org/articles/1_3_Stockwell.pdf), accessed March, 21 2013.

Using English for Academic Purpose (UEFAP). *Features of Academic Writing*. (Online), (Http://www.uefap..com/writing/feature).

Using English for Academic Purpose (UEFAP). Genres in academic writing: Writing Introductions. (Online), (Http://www.uefap..com/writing/feature).

- 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*, (Online) 2 (1): 65-77, (http://teslcanadajournal.ca/index.php/tesl/article), accessed February 23, 2013.
- Yu, G. 2007. Lexical Diversity in MELAB Writing and Speaking Task Performances. Graduate School of Education. University of Bristol, UK. (Online), (http://www.cambridgemichigan.org/sites/default/files/resources/SpaanP apers/Spaan_V5_YuG.pdf), accessed March 21, 2013.
- Salkind, N.J. 2000. *Statistics for People Who Think They Hate Statistics*. California: Sage Publication, Inc.

CURRICULUM VITAE



Ratna Dewi was born in Makassar, South Sulawesi on February 23, 1966. She is the oldest with one sister and one brother from a couple of teachers, St. Rostinah and Ibrahim Mannassai. She graduated from SDN Labuang Baji Makassar in 1977, SMPN 6 Makassar in 1981, and SMAN I Makassar in 1984.

At the beginning, she was not interested to be a teacher, but when she was involved in KKN, the faculty assigned her to teach in a senior high school. Since that

time, she felt that being a teacher was a job that she loved to. After that she became a teaher at Muallimin Muhammadiyah Cabang Makassar and SMA Muhammadiyah Tallo in 1987-1993.

She graduated from English Department, Hasanuddin University in 1993. Her master degree was obtained in 2002. At the same year with the completion of her master degree, she began to teach at Muhammadiyah University of Makassar. Then in 2009 she was awarded BPPS scholarship from Indonesian Directorate of Higher Education to study Doctorate Program in ELT. Her writings are *"Communication Strategies Used by the English Education Department Students of Unismuh Makassar in Speaking English"* (2002); *"English Needs Analysis of PGSD Students in Relation to English as a Local Content Subject at Elementary Schools in Makassar"* (2008); *"Improving the Students' Ability in Writing Paragraph Using Writing Process and Self-Evaluation through Cooperative Learning"* (2008); *"Analyzing the English Academic Needs of the Lecturers of Muhammadiyah University of Makassar"* (2009); *"An English Course for Unismuh Lecturers"* (2010) and *"Encouraging Students to be Autonomous Learners in Studying English at Senior High School"* (2011).