

## 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 firstly tagged in Stanford POS Tagger then lemmatized using MORPHA. Next, the output 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, CTR, 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 WECCCL and the values of argumentative 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 undergraduate 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 desain 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>. Artikel 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, CTRR, 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 frase (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

ix

dalam artikel hasil penelitian dibandingkan 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 WECCCL 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 memperoleh 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 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 lexical 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 frase 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.