LEXICAL COMPLEXITY IN THE INTRODUCTIONS OF UNDERGRADUATE STUDENTS' RESEARCH ARTICLES

Ratna Dewi
Prodi Pendidikan Bahasa Inggris, FKIP Universitas Muhammadiyah Makassar
Jalan Sultan Alauddin No. 259, Makassar 90223, Indonesia
ratnadewi@unsmuh.ac.id

ABSTRACT

Lexical complexity refers to the various different words employed in the introductions of the undergraduate students' research articles. The implementation of lexical complexity describes the writers' overall development of lexical complexity use in the target language. This study aims to find out the employment of lexical complexity in the introductions of undergraduate students' research articles. It utilized a quantitative design through corpus based analysis.The corpus studied consisted of 134 introductions of undergraduate students' research articles. The data were analyzed using Web-based Lexical Complexity Analyzer. Then to decide the level of complexity applied in the students' articles, the results were compared to the lexical complexity of Chinese learners' oral narratives. The findings show that lexical density, lexical sophistication, and lexical variation are high except for verb variation, noun variation, adjective variation, adverb variation and modifier variation (cf. Lu, 2010, 2012; Ai & Lu, 2013). It is expected that this article could provide information about the lexical complexity which is needed to be improved in the undergraduate students' research articles.

Key words: *lexical complexity, research articles, undergraduate students*

Lexical complexity characterizes academic written texts of advanced writers. High proficient writers emerge with the more sophisticated vocabularies, those are Base Word 3, the University Word List (UWL) and 'not in any list words (Laufer & Nation, 1995). Lexical complexity use describes the writer' ability to communicate effectively in written form (Lu, 2012) (Ai & Lu, 2010)(Lu, 2010; Lu, 2012; Ai & Lu, 2010). Due to the fact, the existence of lexical complexity in students' academic texts sets forth the students' writing proficiency. Therefore, lexical complexity proficiency in writing academic texts such as research articles is undoubtedly required.

The appearance of lexical complexity in academic text is also the nature of the text itself that loads complex ideas, which need lexical complexity 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 Base Word 3, in 'not in any list and University Word List. Moreover, complex ideas are commonly written in complex lexis in order to accommodate the needs for describing and explaining specification. Pertaining to the nature of academic text, a writer, in general, needs to implement lexical complexity in their academic texts.

In short, academic texts are characterized by the extensive use of lexical 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.

So far, the studies done are mainly focused on the differences of the existence of lexical complexity 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 complexity 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 Base Word 3, 'Not-in-any-lists' words, and the UWL.

Other research related to Test of Written English explains that lexical complexity is 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 are given based on the extent of word type used in the text, the intensive use of advanced or derived words (unique and longer words) and the proportion of content words exhibited in the text. The words employed by the writers in their writing describe their lexical complexity which is the part of language criteria that reflected the writer's proficiency.

Different research related to lexical complexity is 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 variety. At that time, the learners became more challenged to perform their capacity to use more advanced language.

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

In the Indonesian context, the only research done to the written texts was to study the lexical richness or in this study said as lexical sophistication. Afini and Cahyono (2012) found that both male and female students used the 2,000 most frequent words repetitively. In other words, the students' lexical sophistication was considered low since 79.12 % of the word families used were included as high frequency words.

Considering the vacuity of research on lexical complexity, especially the employment of lexical complexity in academic texts, this present research aims to analyze the lexical complexity employed in the introductions of undergraduate students' research articles as evidence for their acquisition after learning in English Department of Faculty of Letters, Malang State University. This problem is specifically answered by finding the values got by the students in the employment of lexical complexity covering lexical density, lexical sophistication, and lexical variation.

METHOD

Employing a quantitative design through corpus based analysis, this study examined the employment of lexical complexity in the introduction of undergraduate students' research articles using Web-based Lexical Complexity Analyzer - Batch Mode which was accessed in http://aihaiyang.com/software/lca/batch/. The corpus studied was the undergraduate students' research articles published on line at http://jurnal-online.um.ac.id/article/7. They comprised 134 research articles of the English Department alumni of State University of Malang in 2012 and 2013.

Lexical Complexity Analyzer (LCA) required articles in the form of file .txt. in which the content had been in the forms of only paragraphs without pictures, graphs, tables, figures, references, title and subtitles. These data had also been ascertained following the American Spelling through the process of scanning. The output of Web-based Lexical Complexity Analyzer was in the form of numbers which described the values of each criteria of lexical complexity, namely, lexical density, lexical sophistication, and lexical variation. Lexical density was anlalyzed using lexical density measure (LD), lexical sophistication using (LS1, LS2, VS1, VS2, CVS1), lexical variation using (NDW, NDW-50, NDW-ER, NDW-ES, TTR, MSTTR, CTTR, RTTR, AdjV, AdvV, ModV).

Since all the outputs from each measurers were in the form of numbers, in order to know the level of complexity of the lexis, the values were compared to the values of lexical complexity of Chinese learners' spoken narratives which were analyzed using similar software.

FINDINGS AND DISCUSSION

The Undergraduate Students' Lexical Complexity

The Lexical Complexity Analyzer (LCA) developed by Professor Xiaofei Lu and Ai Haiyang at the Pennsylvania State University, is a tool that allows language teachers and researchers to analyze the lexical complexity of written English language samples, using 25 different measures of lexical density, variation and sophistication proposed in the first and second language development literature (Ai: 2016). Notably, the LCA the LCA 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 researcsher compares the count results of the lexical complexity of the introductions of undergraduate students' research articles and the lexical complexity of Chinese learners' oral narratives.

Lexical Density in Undergraduate Students' Research Articles

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

Tabel 1. The Undergraduate Students' Lexical Density

	Lexical Density Values
Mean Value	0.53
Maximum Value	0.59
Minimum Value	0.48

Table 1 describes the density of the lexical words identified in the introductions of undergraduate students' research articles. It informs that the density mean of the lexical words used by the undergraduate students is 0.53 of the total number of words used in the research articles. the representation of the employment of lexical density in the undergraduate students' research articles shown by the mean value obtained are 0.53. This value is higher than that of Chinese oral narratives (0.414). In other words, the number of lexical words implemented in the introductions of undergraduate students' research articles are bigger than those in Chinese oral narratives.

Lexical Sophistiction in Undergraduate Students' Research Articles

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

Table 2. The Undergraduate Students' Lexical Sophistication

	Lexical Sophistication Values				
	LS1	LS2	VS1	VS2	CVS1
Mean Values	0.30	0.28	0.11	1.04	0.66
Maximum Values	0.45	0.39	0.39	5.44	1.65
Minimum Values	0.18	0.14	0.00	0.00	0.00

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 (0.30, 0.28, & 0.11) are higher compared with the mean values of Chinese Learners' spoken narratives (0.23, 0.26, & 0.07). Meanwhile through the mean value of lexical sophistication using transformed measures such as VS2 (1.04) and CVS1 (0.66), the number of advanced words used by the introductions of undergraduate students are bigger than those of Chinese learners' spoken narratives (0.31 & 0.33). These results are in line with the suggestions recommended by (Laufer, 1994; Linnarud 1986) that there are different roles of lexical sophistication played in spoken and written proficiency.

Lexical Variation in Undergraduate Students' Research Articles

Lexical variation of the words employed in the undergraduate students' research articles were identified based on the number of different words, type token ratio, verb diversity, and lexical word diversity. Tabel 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).

Tabel 3. The Undergraduate Students' Number of Different Words

	Lexical Variation Values			
	Number of Different Words (NDW)			
	NDW	NDW-50	NDW-ER50	NDW-ES50
Mean Values	264.77	37.38	38.74	37.42
Maximum Values	465	45	42.30	42.10
Minimum Values	70	30	32.90	31.20

NDW measure counted number of different words or number of word types in a text. NDW-50 calculated number of diffrent word types in the first fifty words of sample. NDW-ER50 computed the mean of the number of different word of 10 random 50-word samples. NDW-ES50 accounted for the mean of word types of 10 random 50-word sequences. Table 3 consists of number of different word values of the undergraduate students' research articles in four measures. Counting the undergraduate 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 second way to identify lexical variation of the words employed in the research articles is using type per token ratio (TTR). Tabel 4 contains values which represent the counts of lexical variation in the research articles using six measures, they are TTR, MSTTR, CTTR, RTTR, LogTTR, and UBER. TTR calculate the number of word types to the number of tokens in the research articles. MSTTR divides a sample into successive segments of a given length and then calcultate the average TTR of all segments. MSTTR, CTTR, RTTR, LogTTR, and UBER are TTR transformation with different formula implemented in counting lexical variation in the research articles. The results of TTR analysis using these six measures are found in Table 4. Compared with the type token ratio of Chinese English learners (0.686 & 4.942), the introductions of undergraduate students' type token ratio is higher which means the undergraduate students

employ number of different word types which are higher than those of Chinese learners.

Tabel 4. The Undergraduate Students' Type Token Ratio

	Lexical Variation Values					
	Type Token Ratio					
	TTR	MSTTR	CTTR	RTTR	LogTTR	UBER
Mean Values	0.38	0.75	6.95	9.83	0.85	19.16
Maximum Values	0.65	0.82	9.09	12.86	0.91	26.65
Minimum Values	0.24	0.67	4.66	6.60	0.80	14.88

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 5. The mean values got from these measures are 49.73 and 4.92. These values are higher compared with the mean values of Chinese learners (13.415 & 2.556).

Tabel 5. The Undergraduate Students' Verb Diversity

	Lexical Variation Values		
	Verb Diversity		
	VV1	SVV1	CVV1
Mean Values	31.64	3.93	0.54
Maximum Values	75.94	6.16	0.86
Minimum Values	8.65	2.08	0.35

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) countes the number of verb type to the total number of lexical

wordss. Noun Variation (NV) accounts for the number of noun type to the total number of lexical words. Adjective Variation (AdjV) reckoned on the number of adjective type to the total number of lexical words. Adverb Variation (AdvV) computes the number of adverb type to the total number of lexical word. Modifier Variation (ModV) calculated the number of adjective and adverb types to the total number of lexical words. The count results of lexical word diversity using six measures: LV, VV2, NV, AdjV, AdvV, and ModV are found in Table 6.

Tabel 6. The Undergraduate Students' Lexical Word Diversity

	Lexical Variation Counts					
	Lexical Word Diversity					
	LV	VV2	NV	Adj.V	Adv.V	Mod.V
Mean Values	0.63	0.14	0.48	0.10	0.12	0.15
Maximum Values	1.00	0.21	0.80	0.20	0.05	0.28
Minimum Values	0.36	0.09	0.28	0.06	0.02	0.09

The mean values of lexical word diversity of the undergraduate students' articles and LV, VV2, NV, AdjV, AdvV, and ModV are 0.63, 0.14, 0.48, 0.10, 0.12, and 0.15. Compared with the mean values of Chinese learners using the same measures, they are 0.57, 0.19, 0.59, 0.11, 0.04, and 0.15. The counts explain that the introductions employ fewer verb variation, noun variation and adjective variation.

CONCLUSIONS AND SUGGESTIONS

This study investigates the use of lexical complexity in the introductions of undergraduate students' research articles. The lexical variation based on the number of different words employed, type token ratio and verb diversity are also high (cf. Lu, 2012). However, the lexical word diversities including verb variation, noun variation, adjective variation are low (cf. Lu, 2012).

Considering the implementation of lexical complexity which is contributed to the elegant style of an academic text and to the equivalent quality on the employment of lexical complexity 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.

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.

Brown, H.D. 2007(3rd ed.). *Teaching by Principles: an Interactive Approach to Language Pedagogy*. New York: San Fransisco State University. (http://moodle.adaptland.it/pluginfile.php/4930/mod_resource/content/1/329 WP_08_71.pdf), accessed March 23, 2013.

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.

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.

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...*)

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. 2012. The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*. 96(2): 190-208.

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).

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-and-literature/elal_lswp_vol_4_siskova.pdf), accessed 13/04/2013.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. 1998. *Second Language Development in Writing: Measurers of Fluency, Accuracy, & Complexity*. Honolulu: University of Hawaii Press.