KLAUSA: **Kajian Linguistik, Pembelajaran Bahasa, dan Sastra**
Journal of Linguistics, Literature, and Language Teaching
Universitas Ma Chung
Vol. 8, No. 1, 2024 ISSN: 2620-3804 (e)/ 2301-4822 (p)

LEXICAL DENSITY AND READABILITY IN STUDENTS' THESIS INTRODUCTIONS

Maysarah Siregar[1], Nina Tiya Hara[2], Tengku Silvana Sinar[3]
Maysarahsiregar201999@gmail.com[1], Ninatiyahara@gmail.com[2],
tengkusilvana@usu.ac.id[3]
[1,2,3]Magister of English, University Sumatera Utara

Abstract: This study investigates lexical density and readability in students' undergraduate thesis introductions. Documentation technique is used to collect the data while qualitative content analysis was employed as the method to present the data. The data of this study were content words and sentences taken from the data sources of 20 undergraduate thesis introductions. The scores of lexical density and readability were obtained by using Flesch Reading Ease Tool online. The findings show that all introductions exhibit high lexical density, with scores ranging from 51,28% to 68,08%. Whilst, the readability scores range from 10.92 to 17.58, and they are all categorized as being very difficult writings. In summary, the findings demonstrate that all thesis introductions exhibit a high level of complexity in their writing. These introductions showcase the deliberate use of rich and precise content words. Consequently, they pose a significant challenge for comprehension due to their density.

Keywords: *lexical density, readability, thesis writing*

Received: April, 6,2024 Accepted: May, 12, 2024 Published: June, 24, 2024

I. INTRODUCTION

A thesis introduction is the opening section provides an overview of the research topic as well as to set the tone for the entire research. It serves as a roadmap for the reader, guiding them through the content and helping them understand the significance of the research. The introduction is important due to four reasons. First, introduction describes why the research is conducted. Second, it helps readers to understand the objectives and problems of the study. Third, it is crucial for the writer to prevent assumptions. Last, it explains the hypothesis of the study (Noorzan & Page, 2012). Research shows that writing a research introduction itself is the hardest step for either native speakers or non-native speakers (Swales & Feak, 2012). Novice researchers have often found it difficult to write a thesis introduction as it involves several overlapping rhetorical strategies (Zainuddin & Shaari, 2021).

To produce a good introduction, it is necessary for the students to know how to organize ideas and facts in their writing as well as to understand the characteristics of written language. Halliday and Matthiessen (2004) suggest that written language typically becomes complex by

being lexically dense: It packs a large number of lexical items into each clause. Generally, written texts have higher lexical density than spoken texts because they have a higher proportion of content words than spoken texts (Stubbs, 2002; Johansson, 2008). Lexical density shows the complexity of words within text which describes the development of lexical in the written language (Gultom and Pintubatu, 2021). Lexical density is closely associated with the notion of information packaging as content words in a text; thus, texts with a higher proportion of content words are considered to be dense as they package more information as opposed to texts that have a higher proportion of function words (Johansson, 2009). Since academic texts are informationally and lexically dense, it can be assumed that high-quality academic essays written by students may exhibit higher percentage of content words. A higher lexical density suggests that the student is able to use language in a sophisticated manner, which is characteristic of more advanced writing skills. Lexical density can be used as one of the criteria for evaluating the language proficiency of the students since it can indicate a stronger command of vocabulary and language structure, which are important skills for academic writing.

Eggins (2004:97) states that the lexical density of a text can be calculated by expressing the number of contents carrying words in a text as a proportion of all the worlds in the text. The measurement of lexical density originally was proposed by Ure (1971) that lexical density should be treated as the proportion of the number of lexical words per number of running words. A large majority of the spoken texts has a lexical density under 40%, while a large majority of the written texts has a lexical density is about 40% or higher. It means that if the result of a text is above 40%, it can be stated that it has higher lexical density, and if the text is highly dense, it contains more information which may influence readers to comprehend the text.

Readability is more focused on measuring text understanding (Kate et al., 2010). Bailin (2016) states that readability is concerned with the degree to which it is easy or difficult to understand what is being communicated through written text. The relationship between lexical density and readability is generally inverse. Texts with higher lexical density tend to be more complex and thus have lower readability. This is because an increase in content words often correlates with longer and more complex sentences, which can pose comprehension challenges for readers, especially those with limited vocabulary or language proficiency. Conversely, texts with lower lexical density, containing more function words, tend to be simpler and have higher readability scores. However, it's important to note that while lexical density can influence readability, it's not the sole determinant. Readability depends on many factors, including (a) the average length of sentences in a passage (b) the number of new words a passage contains (c) the grammatical complexity of the language used (Richards & Schmidt, 2002:442).

Thesis are still considered one of the most problematic academic texts both for EFL and ESL students (Wuttisrisiriporn, 2017). Main interest on this academic writing resulting several studies to undergo investigation on lexical density and readability in students' thesis. Regarding previous research on lexical density, various texts have been intensively examined and different outcomes have been achieved. Some investigations such as on introduction (Syarif & Putri, 2018; Mayangsari, 2021), on abstract (Hanafiah & Yusuf, 2016; Nasseri & Thompson, 2021; Susoy, 2023), on findings (Sinar et al., 2023), on results and discussions (Ebrahimi and Heng, 2018) have contributed to show the analysis on lexical density and readability of their writings. Therefore, this study aims to investigate how undergraduate students compose their thesis writing in term of lexical density and readability. This study attempts to unravel patterns in lexical choices and readability level across different thesis

introductions. By achieving these objectives, the study contributes to give insight of how students navigate the challenges of composing thesis introductions.

II. METHOD

The researchers carry out this study to investigate the numbers of lexical density and the level of readability of undergraduate students' thesis introduction. To collect the data, documentation technique is used. Documentation is the technique which is used to collect data and information about related needs in analysis of written text. To present the data, this study applied qualitative content analysis method. Singh in Hanafiah and Yusuf (2021) states that content analysis, sometimes known as document analysis, deals with the systematic examination of current records or documents as sources of data.

2.1 Data Source

The source of the data was 20 introductions of undergraduate thesis from English Literature Department of Hasanuddin University published in 2023. The researchers choose this source because they are all open access, that can be accessed from Hasanuddin university's repository at their website www.repository.unhas.ac.id. These theses are selected randomly to ensure diversity and to represent a subset of the overall theses available. By choosing them randomly, researchers avoid bias and capture a broader range of content.

2.2 Technique of Analyzing the Data

In analyzing lexical density, a web-based measurement named "Analyzed My Writing" is utilized from its online website at www.analyzemywriting.com. This tool works by counting the number of lexical items which is divided by the total number of words, then the result is multiplied by 100, which is formulated as follows:

$$Lexical\ density\ percentage: \frac{number\ of\ content\ words}{total\ number\ of\ words}\ x\ 100$$

Source: (Halliday, 1985)

Whereas, in terms of readability, Flesch Kincaid reading ease formula is used as proposed by Ure (1971). The Flesch Reading Ease Readability tool calculates the reading ease score. It works based on the following formula:

RE= 206.835 – (1.015 x ASL) – (84.6 x

Source: (Sinar, 2023)

1) RE is Readability Ease

2) The number of words was divided by the number of sentences to collect the number of sentence lengths (ASL).

3) The number of syllables was divided by the number of words to get the average number of syllables per word (ASW).

4) The result matches the Flesch Reading Ease Score indicator on a scale from 1 to 100. The higher the score, the easier the text is to read, and conversely, the lesser the score, the more difficult the text is to read. The index of readability is described as follows:

Flesch reading ease	Description of style	Educational Attainment Level (US)
0-30	Very difficult	Postgraduate
30-50	Difficult	Undergraduate
50-60	Fairly difficult	Grade 10-12
60-70	Standard	Grade 8-9
70-80	Fairly easy	Grade 7
80-90	Easy	Grade 6
90-100	Very easy	Grade 5

Source: (Courtis & Hassan, 2002)

III. FINDINGS AND DISCUSSION

3. 1 Lexical Density

The result of content word calculation, lexical density (Ure, 1971), is calculated using analysis parameters. The texts meet the written language criteria if the content words result is more than 40%. The table shows the results of this study, which measured the lexical density of 20 datasets of students' thesis introduction. The total words obtained in the 20 texts' introductions ranged from 194 to 1948. LD scores in these texts are significantly above 50. It means the results prove that all texts have high lexical density, above the average number of 40%. The results are shown in Table 1:

No.	Noun	Adjectives	verb	Adverb	Total words	Lexical density score
1	31.38%	7.02%	12.02%	4.35%	1539	54.78%
2	28.48%	8.2%	13.36%	2.76%	1085	52.81%
3	29.79%	7.2%	10.26%	4.03%	819	51.28%
4	29.52%	6.72%	13.3%	4%	1948	53.54%
5	32.82%	9.4%	14.15%	3.72%	1639	60.1%
6	34.62%	7.69%	12.5%	1.92%	832	56.73%
7	32.15%	3.98%	10.83%	4.44%	563	55.42%
8	54.87%	2.51%	2.51%	3.57%	1037	68.08%
9	31.36%	8.32%	15.47%	2.93%	1076	58.09%
10	33.08%	7.21%	13.46%	5%	1040	58.75%
11	32.91%	7.94%	10.7%	3.11%	869	54.66%
12	33.56%	5.78%	11.33%	0.89%	450	51.56%

13	29.9%	8.71%	13.5%	4.5%	689	56.6%
14	32.8%	7.91%	10.67%	3.1%	872	54.47%
15	33.56%	5.78%	11.33%	0.89%	450	51.56%
16	34.84%	7.41%	11.46%	4.98%	864	58.68%
17	56.73%	31.33%	10.2%	9.99%	1382	56.73%
18	29.38%	9.79%	13.92%	7.22%	194	60.31%
19	27.39%	7.43%	17.62%	4.07%	982	56.52%
20	31.15%	7.21%	14.29%	3.6%	777	56.24%
Average						56.165%

Table 1. *Lexical density score of Indonesian students' thesis Introductions*

In relation to table 1, lexical density in written text refers to the proportion of content words (such as nouns, adjectives, verbs, and adverbs) to the total number of words. Lexical density serves as a metric indicating the informativeness or content richness of a text. Higher lexical density texts contain more content words and are likely to convey more information. These texts exhibit richer vocabulary due to the prevalence of content words. On the contrary, lower lexical density texts have higher proportion of function words (such as articles, prepositions, and conjunctions). While these texts may be easier to read, they tend to be less informative. The lexical density scores in Table 1 range from 51.28% to 68.08%. This indicates a moderate to high amount of lexical words in the texts. Text No. 8 has the maximum lexical density (68.08%), making it the most dense and complicated text. Conversely, Text No. 3 has the lowest lexical density (51.28%), making it the least dense and less complex. The varying percentages of distinct lexical word classes (nouns, adjectives, verbs, and adverbs) provide insight into each text's characteristics and style (e.g., descriptive, narrative, analytical).

3.2 Readability Score

Table 2 presents the total readability scores of 20 thesis introductions. These scores are considered significant outcomes of text assessment as they can provide insights into how easily a text can be understood by its audience. Higher scores indicate more complex or challenging texts. Readability scores matter because they impact how well readers engage with the content. Instructors use them to assess student writing, and readers consider them when evaluating texts. These scores are based on Courtis & Hassan's (2002) readability index, which determines the difficulty level of the text based on linguistic features. Notably, all 20 thesis introductions exhibit a single style, which is very difficult. These introductions are typically written by postgraduate students, suggesting a common level of complexity. For specific readability scores, see Table 2:

o.	otal words	Total sentence	Total syllables	ASL	Readability score	Interpretation
.	539	61	2478	25.13	14.35	very difficuilt

.	085	46	1809	23.59	14.39	very difficuilt
.	19	31	1296	26.35	14.5	very difficuilt
.	948	103	3462	18.68	14.51	very difficuilt
.	639	64	3024	25.48	17.58	very difficuilt
.	32	34	1291	23	13.42	very difficuilt
.	63	38	923	14.79	10.92	very difficuilt
.	037	57	2078	17.65	16.84	very difficuilt
.	706	137	3061	12.42	12.55	very difficuilt
0.	040	55	1809	18.82	13.25	very difficuilt
1.	69	33	1501	26.3	15.06	very difficuilt
2.	50	16	750	26.31	15.05	very difficuilt
3.	89	30	1149	22.87	13.05	very difficuilt
4.	72	33	1504	26.39	15.07	very difficuilt
5.	50	17	750	26.12	14.4	very difficuilt
6.	64	38	1410	22.71	12.53	very difficuilt
7.	382	61	2462	22.59	14.27	very difficuilt
8.	94	366	8	24.25	16.13	very difficuilt
9.	82	45	1648	20.96	12.72	very difficuilt
0.	77	47	1418	15.98	12.39	very difficuilt

Table 2. *Total readability scores of Indonesia students' thesis introduction*

The readability score is a measure that indicates the difficulty level of a text, based on factors such as sentence length, word length, and other linguistic characteristics. Higher readability scores generally correspond to more difficult, complex texts that require a higher level of education or reading ability to comprehend. In reference to the table 2, The readability scores range from 10.92 to 17.58, with the majority falling between 13.25 and 14.51. The lowest readability score is 10.92, which corresponds to No. 7. This suggests the text is the most readable and accessible, likely requiring a lower level of education or reading ability to understand. The highest readability score is 17.58, found in text No. 5. This indicates that the text is the most challenging, potentially requiring a higher level of education or reading proficiency to comprehend. The majority of the texts have readability scores ranging from 13-15, which is considered to be at the upper end of the "average" difficulty level, which is suitable for a general audience with high education. The readability measurements in each student's

writing demonstrate familiarity with English structure. Students at Hasanuddin University can use content terms to build sophisticated thesis introductions. They exhibit a postgraduate level of lexical density, which aligns with the expectations for advanced academic writing. They can demonstrate sophisticated language use and a strong command of vocabulary. In summary, this categorization suggests that the student's writing meets the standards for postgraduate-level lexical complexity.

In line with this result, a study conducted by Mayangsari (2021) also found similar outcomes. The study shows that readability result of the selected introduction sections are mostly categorized into difficult texts. It turned out that most of them composed lexically dense in which most of the lexical density level are above 0.50. This means that the introduction sections are composed with much information address to the readers in terms of background of the research and literature reviews. The study suggests that higher lexical density offers more information and more difficult text to comprehend. In contrast, Syarif & Putri (2018) found that undergraduate students thesis introductions are less dense. Giving the facts that the texts do not contain characteristics of good academic writings reveals that the students still have limited knowledge about the language use in writing academic texts.

IV. CONCLUSION

Based on the analysis of the data provided, several important conclusions can be drawn regarding the readability and lexical density of Hasanuddin University students' thesis. Lexical density scores range from 51.28% to 68.08%, indicating the texts have moderate to high level of lexical density. The text with the highest lexical density is text No. 8 (68.08%), in contrast, the text with the lowest lexical density is text No. 3 (51.28%), making it the less complex text among others. The percentage of different lexical word classes (nouns, adjectives, verbs, adverbs) varies across texts, thus providing insight into the characteristics and writing style of each text. Regarding readability, scores range from 10.92 to 17.58, with the lowest readability score is10.92, while the text with the highest readability score is 17.58 indicating it as the most challenging text. Most texts have readability scores from 13 to 15, which are all considered "difficult" and suitable for high education readers. Average sentence length also varies, from 12.42 words per sentence to 26.35 words per sentence, which further contributes to varying levels of readability. Overall, the data shows that this thesis have moderate to high level lexical density, with very difficult readability. High lexical density in students' writing is a signal of advanced writing abilities and suggests that they possess the linguistic skills necessary to excel in academic contexts. It reflects not only a strong command of vocabulary but also a capacity for critical thinking, effective communication, and sophisticated expression of ideas.

REFERENCES

Bailin, Alan., & Ann, Grafsten. (2016). *Readability: Text and Context*. London: Palgrave Macmillan.

Courtis, J. K., & Hassan, S. (2002). *Reading Ease of Bilingual Annual Reports*. *Journal of Business Communication 39*(4), 394 - 413.

DuBay, W. H. (2004). *The Principles of Readability*. Impact Information.

Ebrahimi, S. F., & Heng, C. S. (2018). *Lexical Complexity in Master's Dissertations*. GEMA Online® Journal of Language Studies, 18(1), 33-49.

Eggins, Suzanne. 2004. *An Introduction to Systemic Functional Linguistics Second Language.* New York: Continuum International Publishing Group.

Flesch, Rudolf. (1949). *The Art of Readable Writing*. New York: Harper & Row Publisher.

Halliday, M. A. K. (1985). *Spoken and Written Language.* New York: Oxford University Press.

Hanafiah, Ridwan., and Yusuf, Muhammad. (2016). *Lexical Density and Grammatical Intricacy in Linguistic Thesis Abstract: A Qualitative Content Analysis*. Proceedings of the 1st English Education International Conference (EEIC) in conjunction with the 2nd Reciprocal Graduate Research Symposium (RGRS) of the Consortium of Asia-Pacific Education Universities (CAPEU) between Sultan Idris Education University and Syiah Kuala University November 12-13.

Johansson, Victoria. (2008). *Lexical diversity and lexical density in speech and writing: a developmental perspective.* Lund University, Dept. of Linguistics and Phonetics Working Papers 53 (2008), 61.

Johansson, V. (2009). *Lexical diversity and lexical density in speech and writing: A developmental perspective*. Lund Working Papers in Linguistics, 53, 61–79.

Kate, R., Luo, X., Patwardhan, S., Franz, M., Florian, R., Mooney, R., Roukos, S., & Welty, C (2010). *Learning to Predict Readability Using Diverse Linguistic Features*. In: Proceedings of the 23rd International Conference on Computational Linguistics, (pp. 546–554). Association for Computational Linguistics.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.* Naval Technical Training Command Millington TN Research Branch.

Mayangsari, A., Wuli Fitriati, S., & Sutopo, D. (2021). *Lexical Complexity and Readibility Realized in The Introduction Sections of Selected English Journals. English Education Journal*, *11*(2), 298-307. Retrieved from https://doi.org/10.15294/eej.v11i1.44074.

Nasseri, Maryam., & Thompson., Paul. (2021). *Lexical Density and Diversity in Dissertation Abstracts: Revisiting English L1 vs. L2 Text Differences*. Journal of Assessing Writing Volume 47, January 2021. Retrieved from https://doi.org/10.1016/j.asw.2020.100511.

Noorzan, M. N. & Page, G. M. (2012). *Writing Your Thesis*. Kuala Lumpur: Pearson Malaysia.

Sinar, T. S., Zein, T. T., Ganie, R., Syarfina, T., Mahriyuni, Yusuf, M., & Rangkuti, R. (2023). *Content Words and Readability in Students' Thesis Findings*. Journal of Curriculum and Teaching, 12(6), 347-355.

Stubbs. M. (2002). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing.

Susoy, Zafer. *Lexical Density, Lexical Diversity and Academic Vocabulary Use: Differences in Dissertation Abstracts.* Acuity: Journal of English Language Pedagogy, Literature and Culture, v8 n2 p198-210 2023.

Swales, J. M., & Feak, C. B. (2012). *Academic Writing for Graduate Students: Essential Tasks and Skills (3rd ed.)*. Ann Arbor, MI: The University of Michigan Press.

Swales, J. M. (1990). *Genre Analysis*. Cambridge: Cambridge University Press.

Syafitri, S.E., Sinar, T. S., Mulyadi., & Masdiana. (2022). *Types of Minor Clauses in Kindergarten English Interactions.* World Journal of English Language,12(04).

Syarif, H., & Putri, R. E. (2018). *How Lexical Density Reveals Students' Ability in Writing Academic Text*. LINGUA DIDAKTIKA, 12(2), 86-94.

Wuttisrisiriporn, Niwat. *Comparative Rhetorical Organization of ELT Thesis Introductions Composed by Thai and American Students*. Language Institute, Burapha University, Chonburi, Thailand.

Zainuddin, S.Z, Shaari, A.H. *A Genre-Inspired Investigation of Establishing the Territory in Thesis Introductions by Malaysian ESL Writers*. 3L: Language, Linguistics, Literature. *The Southeast Asian Journal of English Language Studies Vol 27(2), June 2021* retrieved from http://doi.org/10.17576/3L-2021-2702-11 144.