

Content Words and Readability in Students' Thesis Findings

T. Silvana Sinar^{1*}, T. Thyrhaya Zein¹, Rohani Ganie¹, T. Syarfina¹, Mahriyuni¹, Muhammad Yusuf¹ & Rahmadsyah Rangkuti¹

¹Faculty of Cultural Sciences, Universitas Sumatera Utara, Indonesia

*Correspondence: English Literature Study Program, Universitas Sumatera Utara, Medan, Indonesia. E-mail: tengkusilvana@usu.ac.id

Received: December 17, 2022

Accepted: April 2, 2023

Online Published: November 17, 2023

doi:10.5430/jct.v12n6p347

URL: <https://doi.org/10.5430/jct.v12n6p347>

Abstract

This study investigates the content words and readability in bachelor's thesis findings in the English Literature Program at the University of Sumatera Utara. Qualitative analysis was applied in this study. The data for this study were content words and sentences taken from the data sources of 13 bachelor's thesis findings. The content words were collected using a lexical density online tool, and the data for readability was collected and analyzed using an online Flesch Reading Ease tool. The results show that the lexical density of the content words ranges from 50.47% – 57.5%. Whilst the readability of the 13 texts range from 19.1 – 61.7. The average score of content word density indicates that the theses' findings present concise information as represented in scientific writing, and the readability style ranges from "very difficult to read" to "standard readable". In conclusion, these findings can be categorized as densely written language and content words, supported by college students' increasingly intricate choice of words and sentences frequently read.

Keywords: content words, lexical density, readability, thesis writing

1. Introduction

The thesis findings significantly relate to academic writing in university data sets. It presents the data analysis results, forming a prominent part of the paper. Lecturers acknowledge that the finding should be suitable for the problems of the study. The findings are combined with a bachelor's thesis discussion, although it is only one or a half pages long.

Although extensive guidance to the students has been given, determining acceptable findings has proven difficult. Many undergraduate students in the English Literature department have been continually writing the analysis in the finding section and have yet to explain why the results are what they are. Moreover, the students must write a concise argument for their analysis in the finding section. The core idea is that students should be able to give a reason based on the result of the study. The study can be significantly helpful for students to enhance their ability to write their findings and provide references to supervisors in guiding them in completing their undergraduate thesis.

Recently, lecturers also examined the readability and density of students' theses. The writing becomes more readable when superfluous, repeatable, and long words are removed from the thesis findings. Furthermore, the findings should be dense to provide information on the text's topic.

This study clarifies the concept of *density* by Halliday (2005, p.83) as the proportion of content words (nouns, adjectives, verbs, and adverbs) to the total number of words that makes a text lexically dense. The content words include nouns (tell us the subject), adjectives (tell us more about the subject), verbs (tell us what they do), and adverbs (tell us how they do it).

A spoken text's ideal proportion of lexical densities is under 40%, but higher than 40% in a written language. Likewise, Eggins (2004) further claimed that a written language has a high lexical density, while a spoken language has a low lexical density. The ideal lexical densities in a written language are more significant than 40%.

The use of the readability formula in this study is to measure the difficulty or ease of a text for the readers' comprehensibility of what was written. The idea of readability (Harris & Hodges, 1995) is to examine a text's readability to see how it interacts with the reader's comprehension. To measure readability, we need to use a formula to analyze the text and the readers' understanding. We use Flesch Reading Ease Readability to calculate the variables in writing.

The classic readability research of Flesch and Gunning was studied for several years in 1949, and the results of a 10-year study of the editorial content of several magazines in America were published. Carver (1975-1976) assessed the complexity of texts using a method of *qualified raters*. He claimed this approach was slightly more accurate than the Dale, Chall, and Flesch Reading Ease formulas, catering grade-level scores to grade 18. Also, in 1975, Singer introduced the Singer Eyeball Estimate of Readability method, which he claimed was accurate. Chall and her associates (1996) then published *Qualitative Assessment of Text Difficulty: A Practical Guide for Teachers and Writers*. It used graded passages as “scales” for the levelling of texts. Sociolinguistic research from the early 1980s revealed that literacy as a social process and literacy as a single set of skills could be acquired—or tested—regardless of one’s interests or social, political, and economic environment. According to Cope and Kalantzis (2000), adults practise multiple literacies that transform in different settings and are embedded in visual, auditory, spatial, and other semiotic systems. It is also argued that testing for knowledge (by checklists or other methods) is a more accurate indicator of literacy.

Regarding relevant research on lexical density (LD) and grammatical intricacy, To (2018) examined four EFL textbooks to differentiate between complexity and textbook levels. This study discovered that linguistic complexity increases when the textbook level progresses from elementary to upper intermediate. On the other hand, none of the textbook levels depicts the peak of complexity. A strong relationship between nominalization, grammatical metaphor, and LD can be indicated through the data.

Amelia et al. (2020) investigated the LD and grammatical complexity of debaters’ talks in their study. Debaters’ speeches are reported to have higher LD and grammatical intricacy and ‘can be classified as both written and spoken language. Furthermore, by assessing the language complexity in the introduction section of a bachelor’s thesis, Zein et al. (2020) indicated that this section is categorized as a written language with an LD of 53.54%.

Based on their research findings, Gregori-Signes and Clavel-Arrotian (2015) show that lexical richness across two writing pieces of the same students may be distinguished. For students learning English as a second language, lexical richness measures how vocabulary size is represented in use. In this study, the lexical richness corresponds well with the measure of vocabulary size. This reliable measure of lexical richness is valuable in evaluating the elements that influence writing quality assessments while examining the relationship between vocabulary growth and vocabulary use.

By using Adelex Analyser (ADA) to run the participants’ group proposals, Nguyen and Edwards (2015) found that the average LD ranged from 27.84% to 39.17%, with the mean being 32.29%. In addition, their Year 2 participants wrote 32 content words (nouns, verbs, adjectives, and adverbs) per 100 words on average. Finally, in their study, Zein et al. (2020) demonstrated that the introduction section of Indonesian students’ bachelor’s thesis is classified as a written language with an LD of 53.54%.

For non-native English speakers, LD in reading texts is around 46.3% (To, Fan, & Thomas, 2013) because of the participants’ use of active voice, hence the low LD. According to Gilquin and Paquot (2008), the LD drops when participants use nominalizations less than they should.

Regarding readability, Gazni (2011) used the Flesch reading ease formula to investigate the text reading level of abstracts of articles in the five most cited institutions. The academicians’ 22 subject fields indicate that the abstract needs to have higher readability or be extremely difficult to read. This important finding is beneficial for scientific and educational policymakers and information-seeking researchers, especially in the fields of medicine and chemistry, which have the most complex test scores. In contrast, mathematics and space sciences have higher readability.

This article considers that the readability level should be crucial as students read a text for a specific purpose. DuBay (2007) acknowledges that the achievement of different levels of reading skills depends on education and reading practices. For example, students who graduated from high schools typically reach the 10th-grade level, yet they only achieve the 8th-grade reading level. In the same way, Sticht et al. (1987) and Sticht and Armstrong (1994) admit that advanced readers who possess extensive knowledge perform well in a wide range of domains, while poor readers do not.

Ebrahimi and Heng (2018) analyzed selected articles from prestigious journals in their 'results and discussion sections. As a result, the macro functions realized the choices, frequencies, and discourse functions of grammatical rules in the ‘results and discussions’ parts of the four articles. This finding implies the implementation of the restricted field conventions.

The descriptive findings of readability are critical in examining the student thesis, particularly in the finding section.

Therefore, the problems are formulated as "How are the content words density and readability shown in student thesis findings in the English Literature Program at the University of Sumatera Utara, and why is the result realized in the way it does?"

2. Method

2.1 Data Source and Data

The data source of this study is 13 undergraduate thesis findings from the English Literature department thesis collection published in 2021. From this data source, the content words (nouns, verbs, adjectives, and adverbs) were obtained and then measured using the LD formula of Halliday (1985b) and Ure (1971). The formula for LD calculation is the number of lexical items times 100 divided by the total number of words (Halliday, 1985b; Ure, 1971). Ure proposed that a higher proportion of LD is at least 40%. Afterwards, the Flesch Reading Ease Readability tool was used to analyze the data to determine the reading ease score.

2.2 Data Collection Technique

In a turbulent time, owing to significant changes in information technology, manually following the calculation of LD is no longer necessary: the number of lexical items x 100 divided by the total number of words. In this study, the data collection technique for content words employed the website formula for LD calculation. It identified each word as either a lexical density word or not.

The Flesch Reading Ease Readability tool calculated the reading ease score. It worked based on the following formula: $RE = 206.835 - (1.015 \times ASL) - (84.6 \times ASW)$

- 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 range of 0 – 10 is "complicated", 10 – 30 "very difficult", 30 – 50 being "difficult", 50 – 60 being "fairly difficult", 60 – 70 being "standard", 70 – 80 being "fairly easy", 80 – 90 being "easy" and 90 – 100 being "very easy".

2.3 The Technique of Data Analysis

To get the total number of content words and analysis of LD, the following sentence was selected to exemplify the measurement of lexical words (*italicized*).

The writer found that the majority of the students agree that using learning social platforms is adequate for their improving writing skills, vocabulary and language use.

The sentence above contains 16 lexical words out of 26 total words. The percentages were: nouns 30.77%, adjectives 7.69%, and verbs 23.08%. The content words were divided by the total number of words and lexical items, timed to 100 then divided by the total number of words. The result of density is 61.54%. Hence, this sentence is considered an informative text and categorized as scientific.

In determining the grades for the readability of the text, the technique used to draw the result was as follows:

- 1) Calculated the average number of words divided by the number of sentences to get the ASL.
- 2) Calculated the average number of syllables per word.
- 3) Multiplied the average number of words by 1.015 ($1.015 \times ASL$) and the average number of syllables per word by 84.6 ($84.6 \times AWL$).
- 4) Subtracted 206.835 by ($84.6 \times AWL - (1.015 \times ASL)$).
- 5) Thus, the output ranged from 0 to 100.
- 6) Referring to the conclusion of the text readability score of the sentence below,

"The writer found that most of the students agree that learning using a social platform is effective for improving writing skills, vocabulary and language use".

The sentence above had an average readability level of 15.47. This grade was considered a "very difficult" text and was understandable by an average postgraduate due to its more lexical content.

3. Results and Discussion

3.1 Results

3.1.1 Content Words Density (CWD)

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%. Table 2 shows the results of this study, which measured the content word density of 13 datasets of students' thesis findings. The total words obtained in the 13 texts' findings ranged from 323 to 497. CWD scores in these texts are significantly above 50. The first group receives scores ranging from 57.51% (text 12), 56.47% (text 7) to 55.1% (text 1). The second group ranges from 54.21% (text 11), 53.56% (text 13), 53.52% (text 6), 52.54 % (text 2), 52.36 % (text 5), 52.34% (text 10), 51.14% (text 8), 51.13% (text 9), 51.06% (text 4), and 50.47% (text 3).

It means the results prove that all texts had a high LD, above the average number of 40%, as proposed by Ure (1971). The results are summarised in the following table.

Table 1. Contents Words Density (CWD)

No.	Noun	Verb	Adjective	Adverb	Total Content Words	Total Words	Lexical Density Score
1	31.72%	12.63%	6.72%	4.03%	55.1%	373	55.11%
2	30.96%	11.68%	6.09%	3.81%	52.54%	397	52.54%
3	21.73%	17.52%	8.18%	3.04%	50.47%	429	50.47%
4	28.13%	11.82%	6.15%	4.96%	51.06%	426	51.06%
5	25%	18.63%	5.9%	2.83%	52.36%	428	52.36%
6	29.58%	13.48%	6.24%	4.23%	53.52%	497	53.52%
7	35.93%	9.45%	9.03%	2.05%	56.47%	494	56.47%
8	26.14%	10.23%	6.36%	8.41%	51.14%	440	51.14%
9	28.75%	12.32%	6.16%	3.9%	51.13%	488	51.13%
10	31.54%	10.98%	5.84%	3.97%	52.34%	429	52.34%
11	29.16%	11.85%	10.71%	2.51%	54.23%	443	54.21%
12	30.28%	10.56%	13.89%	2.78%	57.51%	360	57.5%
13	32.2%	10.22%	8.67%	2.48%	53.57%	323	53.56%

Based on the result, the calculation of nouns in 13 texts obtained the highest score among verbs, adjectives and adverbs in all texts. In terms of nouns, most nouns found in the clauses are groups of words with a noun (names of persons, places or things) as their headword and all additional information related to that noun. Halliday (1985; 2004) called nouns the nominal group. It consists of Thing (the noun itself), Deictic (specific, i.e., the, or non-specific, i.e., a or an and post-deictic). A classifier is the closest to the Thing, while Qualifier provides additional defining or circumstantial information about the Thing.

The clauses below are selected from the investigated texts. Most nouns are primarily the result of nominalization, which converts verbs, adjectives, and conjunctions into nouns (things). Take a look at the following extracts taken from the data set.

- 1) Having *analyzed* the *answers* of the *participants* that were *received* from the questionnaires, the *researcher* found some *results* about the *contribution* of *students' speaking skills* towards their *careers* (CWD 46.43%).
- 2) The *findings* showed that *the movie* above often uses *illocutionary acts* such as *representative, directives, commissive, expressive* and *declarative* (CWD 61.9%).
- 3) The *finding* showed that *most* of the *verbal texts* in the *images* show *straightforward statements* with the explicit suggestion that the *verbal texts* are easily *understandable* by the *public* (CWD 55.17%).
- 4) Through the *questionnaire*, it was able to find out how *in-game communication* has *developed* their *English communicative competence*, how they *intend* to *put gaming* in a *fan-engaging way* to *learn to speak English* and how they *plan* to *cope* with their *hardships* in *combating* their *current problem* in *speaking English* (CWD 54.17%)

- 5) Having *discussed* how the *main character, Cora, fights* for her *freedom*, it *concludes* that *Cora chooses* to *run away* from the *plantation* not because she is *afraid* but because she *knows* that she is *powerless* to *counter* herself from *white people* who *take benefit* to *earn so much money* without *feedback* (CWD 52%).
- 6) The *concept of beauty* for *women* is represented by the *slim proportional body shape* with *firm breasts*, *curvy waists* and *plump buttocks* (CWD 63.64%).

In terms of the measurement result, these sentences demonstrate dense information and are considered high-density. According to Ure (1971), the authors write in the style of written discourse suitable to the purpose of thesis findings. As a result, the lexical density increases dramatically when the author employs more nominal, verbal, adverbial, and adjectival groups, such as in the following examples. Regarding grammar, using more nominal groups as the highest total content word score increases the grammatical metaphor. Very few participants involve pronouns as they are used mainly by non-human participants, making the text much more condensed. In the following examples, the result of nominalization again increases the grammatical metaphor in the texts: “the *concept of beauty for women, slim proportional body shape, main characters’ firm breasts, curvy waist and plump buttocks, psychological impacts of mother’s absence, viewer fashion items, the high number of content words, etcetera.*” The authors changed the processes (verbs) and conjunctions into nouns. The percentage is very high compared to verbs, adjectives, and adverbs.

3.1.2 Readability Score

Table 3 shows the total readability scores of these 13 texts, which are believed to be critical results of the text measurement. The scores imply much about the lecturers’ and readers’ judgments of the text’s readability level. They have four styles, i.e., very difficult, fairly difficult, difficult, and standard. The first *very difficult* style is found in Texts 12 and 13. The second style, a *fairly difficult text*, is found in Texts 3, 4, 6, 7, and 9. The readability scores in Texts 1, 2, 10, and 11 are categorized as *difficult*, while Text 8 is a *standard* style.

Very difficult texts are suitable for postgraduate students, *fairly difficult* texts are appropriate for undergraduate or college graduates, and *10th graders easily understand difficult texts*. Meanwhile, *standard* ones are comprehensible for Grades 8 and 9, as they contain plenty of words frequently used daily and are less complex than the previous texts. Furthermore, the measurements indicate that Text 4 belongs to the upper-intermediate level despite having the lowest LD among the four texts. It contradicts Flesh’s (date) method, considering Text 1 (elementary level) appears to be the simplest. Finally, in light of the Educational Attainment Level (US), Text 2 is recommended for undergraduate education, Texts 3 and 4 are appropriate for students in Grades 10–12, whilst Text 1 is suitable for students in Grade 7.

Table 2. Total Readability Scores of the 13 Texts

No. of texts	Total Words	Total Sentences	Total syllables	Average Sentence Length in Words	Readability Score	Description of Style
1	373	15	669	29 or more	30.5	Difficult
2	397	16	692	21	42.1	Difficult
3	429	28	692	25	56.4	Fairly Difficult
4	426	17	655	25	51.9	Fairly Difficult
5	428	16	711	21	44.7	Difficult
6	497	19	772	25	52.4	Fairly Difficult
7	494	23	769	25	52.5	Fairly Difficult
8	440	23	662	25	61.7	Standard
9	488	22	757	17	59.3	Fairly Difficult
10	429	25	718	21	49.2	Difficult
11	443	16	741	21	42.8	Difficult
12	360	15	670	21	29.1	Very Difficult
13	323	7	526	21	19.1	Very Difficult

Having taken seven sentences extracted from the data, sentences 7 to 13 below show the measure of readability level ranging from 11.19 to 26.1, with the sequence being as follows. First, sentence 7 scored 11.19, while Sentence 10 received 14.55. The next group includes Sentence 8 (17.26), Sentence 9 (17.66), Sentence 11 (18.2)

and Sentence 13 (18.38). Finally, only Sentence 12 gets the highest readability level of 26.1.

- 7) The *results show* that the *main characters also serve* as the *narrators of the story Jason Grace Piper Mclean and Leo Valdez*.
- 8) Twenty-nine *effects of emotional neglect and emotional abuse* as the *psychological impacts of a mother's absence*, and there are 16 *psychological impacts found* with 8 *similarities*.
- 9) The *syntactic aspect relates* to the structure of phrases and *sentences related to logical thinking that explains propositions* in a *series of sentences, coherence references* and the *selection of a number of pronouns*.
- 10) This *picture offers* the *represented participants to the viewer fashion items the model is wearing* as the *participant was a specimen in a display case*.
- 11) The *data also shows* that the *researcher successfully packs* the *information* in their *scientific article* by *providing high numbers of content words instead of grammatical items*.
- 12) The *first video does not have verbal language* but *still contains textual language explained by the linguistic analysis*. The *visual analysis is divided* into 16 parts, while the *gestural analysis is contained* in four rows. For *spatial analysis*, the *focus* is on one *scene of the video that contains the locus of attention*.
- 13) On the *other hand*, the *second video also does not embody* the *verbal language* but *presents* the *textual language defined* in the *linguistics analysis section*. For the *visual analysis*, the *scenes are split* into 18 parts.

The readability scores were identified as the number of sentences from the total number of words and then counted by the number of syllables. After counting the number of syllables, the increasingly difficult choice of words, such as *types of illocutionary acts* and *eighty-five utterances used by the main character in the Maleficent movie script*, increases the readability score. Unfortunately, whereas these words are usually familiar to college graduates, high school or junior high school students are less likely to understand them.

Based on the result of their readability measurement in each student's writing above, it can be proven that they are familiar with the English structure. As they write their thesis findings, final-year students can manipulate content words in writing complex texts. They write factual texts with reasoning, putting theories together and supporting them with arguments.

4. Discussion

Table 3. Summary of the Lexical Density and Readability Scores

No.	Lexical Density Score	Readability Score	Description of Style
1	55.11%	30.5	Difficult (Intermediate)
2	52.54%	42.1	Fairly Difficult (Intermediate)
3	50.47%	56.4	Difficult (Intermediate)
4	51.06%	51.9	Difficult (Intermediate)
5	52.36%	44.7	Fairly Difficult (Intermediate)
6	53.52%	52.4	Difficult (Intermediate)
7	56.47%	52,5	Difficult (Intermediate)
8	51.14%	61.7	Difficult (Intermediate)
9	51.13%	59.3	Standard (Intermediate)
10	52.34%	49.2	Fairly Difficult (Intermediate)
11	54.21%	42.8	Fairly Difficult (Intermediate)
12	57.5%	29.1	Very Difficult (Pre-intermediate)
13	53.56%	19.1	Very Difficult (Pre-intermediate)

After selecting the 13 texts of the thesis findings, a measure of the LD of all texts 1 to 13 is correlated with the difficulty level of readability, as seen in Table 4. Text 1 is 55.11%, corresponding to the readability ease of 30.5, while Text 2 has a 52.54% LD which matches the readability level of 42.1. Sentence 3 has an LD of 50.47%, which is relevant to a readability level of 56.4, and Sentence 4 has an LD of 51.06%, corresponding to the readability ease

of 51.9. Likewise, Sentence 5 obtains an LD of 52.36% with a readability level of 44.7. Sentence 6 has an LD of 53.52% and a readability level of 52.4, while the LD of Sentence 7 is 56.47% with a readability level of 52.5.

While Text 8 receives 51.14% and matches 61.7 for the readability ease, Text 9 has 51.13% and obtains the readability level of 59.3. Then, Sentence 10 has an LD of 52.34% and is relevant with a readability of 49.2. Sentence 11 receives an LD of 54.21%. It is matched with the readability ease of 42.8, Sentence 12 has an LD of 57.5% and is relevant with the readability of 29.1, and Sentence 13 obtains an LD of 53.56% and is matched with the readability ease of 19.1.

From the selected sentences of the 13 texts, the measures of LD are as follows. Sentence 7 has an LD of 63.64% and a difficulty level of readability of 11.19. Sentence 8's LD is 52% with a readability level of 17.26, Sentence 9 has an LD of 54.55% and a readability score of 17.66, and Sentence 10's LD is 52% with a readability ease of 14.55. Likewise, Sentence 11 obtains an LD of 65.38% with a readability level of 18.2, Sentence 12 has an LD of 52.63% and a readability ease of 26.1, and an LD of Sentence 13 is 51.28% with a readability level of 18.38.

The student's high ability level is significant between the formula LD and readability relatedness and match. They used various technical words and an indeterminate number of clauses. They can also combine these two skills in writing and reading, resulting in technical vocabulary.

The results show an increase in the total amount of nominalization found in the clauses. This result indicates that the information is organized in nominal groups within clauses. As a result, the texts are read with more condensed information. According to Halliday (2014), nominalization leads to more abstract texts because the subject of the pronoun, such as people, is replaced by abstract participants.

5. Conclusion

The students' writing is considered scientific articles. In other words, the texts' authors constructed their writings using technical terms as scientific writings, and their choices determine the distribution of content words per clause. As a result, nominalization frequently appears throughout the thesis' finding clauses as the authors considerably employ many content words. This phenomenon has shown the importance of nominalization in a scientific text. According to Kazeiman (2013), nominalizations are scientific speech's most common experiential grammatical metaphor. They provide a higher concentration of experiential meaning and a lower occurrence of interpersonal features, such as personal pronouns and modal verbs, resulting in a less personalized presentation of information. Long phrasal compounds and structures in scientific registers are likewise reduced by nominalizations, making them more compact, practical, pragmatic and direct to professionals (Kazeiman, 2013, p.152).

The phenomenon of writing scientific texts has a connection to their readability. The compact and abstract use of nominal groups results in text readability since the clauses are more challenging to read. The authors often changed or modified the verbs or adjectives to create nominalizations.

Moreover, the students' choices of content words find texts increasingly metaphorical, as their congruent texts are written in metaphorical sentences. It demonstrates that their written texts become fairly difficult and extremely difficult at the pre-intermediate level. This study was successful in proving the significance of the measurement. The number of lexical words in each sentence correlates with the degree of readability ease.

References

- Alami, M., Sabbah, M., & Iranmanesh, M. (2013). Male-female discourse difference in terms of lexical density. *Research Journal of Applied Sciences, Engineering and Technology*, 5(23), 5365-5369. <https://doi.org/10.19026/rjaset.5.4202>
- Amelia et al. (2020). Lexical density and grammatical intricacy in debaters' speeches. *Language Literacy: Journal of Linguistics, Literature and Language Teaching*, 4(1), 168-184, June 2020.
- Arizavi, S., Namdari, N., & Mousavi, S. A. (2015). Nominalization in the research article discussion sections of local and international journals of applied linguistics. *Iranian Journal of Applied Linguistics*, 18(1), 1-28.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20.
- Bozdağ, F. Ü. (2014). *Error Mapping and remedial intervention regarding English Tense - Aspect Structures of Turkish EFL Learners* [Doctoral Dissertation]. Republic Of Turkey Çukurova University.
- Carmen, G., & Begoña, C. (2015). Analyzing lexical density and lexical diversity in university students' written

- discourse. *7th International Conference on Corpus Linguistics. Current Work in Corpus Linguistics: Working with Traditionally-conceived Corpora and Beyond (CILC 2015)*.
- Chall, J. S. (1996). American reading achievement: Should we worry?. *Research in the Teaching of English*, 30(3), 303-310.
- Cope, B., & Kalantzis, M. (Eds.). (2000). *Multiliteracies: Literacy learning and the design of social futures*. Psychology Press.
- Du Bay, WH (2007). *Smart Language: Readers, Readability, and the Grading of Text, Impact Information*. Retrieved from <http://www.impact---information.com>
- Eggins, S. (2004). *Introduction to systemic functional linguistics*. A&C Black.
- Gastel, B., & Day, R. A. (2016). *How to write and publish a scientific paper*. Santa Barbara, CA: ABC-CLIO.
- Gavin, B. (2019). *Influence of learners' prior knowledge, L2 proficiency and pre-task planning on L2 lexical complexity*. De Gruyter. <https://doi.org/10.1515/iral-2018-0244>
- Gazni, A. (2011). Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the world. *Journal of Information Science*, 37(3), 273-281.
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41-61.
- Hajar, S., & Sinar, T. S. (2019). "Projection Analysis of Business Text in *The Jakarta Post*" in *The Second Annual International Conference on Language and Literature*, KnE Social Sciences, pages 745-752. <https://doi.org/10.18502/kss.v3i19.4901>
- Halliday, M. A. (2005). On matter and meaning: The two realms of human experience. *Linguistics & the Human Sciences*, 1(1).
- Halliday, M. A. K., & Matthiessen. (2014). *An introduction to functional grammar*. London: Arnold.
- Halliday, M. A. K. (1985). *Spoken and Written Language*. Victoria: Deakin University.
- Halliday, M. A. K. (1989). Some Grammatical Problems in Scientific English in *Australian Review of Applied Linguistics: Genre and Systemic Functional Studies*, 6, 13-37.
- Halliday, M. A. K. (2008). *Complementarities in language*. Beijing: The Commercial Press.
- Harris, T. L., & Hodges, R. E. (1995). *The literacy dictionary: The vocabulary of reading and writing*. International Reading Association
- Hartley, J., Sydes, M., & Blurton, A. (1996). Obtaining information accurately and quickly: are structured abstracts more efficient? *Journal of Information Science*, 22(5), 349-356. <https://doi.org/10.1177/016555159602200503>
- Holtz, M. (2009). Nominalization in scientific discourse. *Proceedings of CL009*, p. 341.
- Johansson, V. (2008). *Lexical diversity and lexical density in speech and writing*. Lund University: Lund University Press.
- Koksal, O. (2013). The impact of the keyword method on vocabulary learning and retention in preparatory french classes in higher education. *International Journal of Academic Research*, 5(5).
- Lorés, R. (2004). On RA abstracts: From rhetorical structure to thematic organization. *English for Specific Purposes*, 23(3), 280-302.
- Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Larivière, V., ... & Zhang, C. (2019). Analyzing linguistic complexity and scientific impact. *Journal of Informetrics*, 13(3), 817-829.
- Ma'mun, N. (2017). Grammatical intricacy in students' writing. *Vision: Journal for Language and Foreign Language Learning*, 6(1), 47-60. <https://doi.org/10.21580/vjv6i11619>
- Nuyen & Edwards. (2015). Language Education in Asia. *Pre-Training Performance*, 6(1).
- Putra, D. A., & Lukmana, I. (2017). Text complexity in senior high school English textbooks: A systemic functional perspective. *Indonesian Journal of Applied Linguistics*, 7(2), 436-444. <https://doi.org/10.17509/ijal.v7i2.8352>
- Ramadhan, G., Santihastuti, A., & Wahjuningsih, E. (2017). The Lexical Density of Erlangga's "English On Sky" Analysis. *Jurnal Edukasi*, 4(3), 38-40.

- To, V. (2018). Linguistic complexity analysis: A case study of commonly-used textbooks in Vietnam. *SAGE Open* (July-September 2018), 1-13. <https://doi.org/10.1177/2158244018787586>
- To, V., Fan, S., & Thomas, D. (2013). Lexical density and readability: A case study of English textbooks. *Internet Journal of Language, Culture and Society*, (37), 61-71.
- Ure, J. (1971). Lexical density and register differentiation. *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics*, 1(1), 443-452.
- Zein et al. (2020). Assessing Complexity in the Introduction Part of Bachelor's Theses: A Systemic Functional Linguistic Framework. *Talent Development & Excellence*, 12(1), 2100-2111.

Acknowledgments

We would like to thank to the Faculty of Cultural Sciences for the permission to conduct this research.

Authors contributions

T. Silvana Sinar, T. Thyrhaya Zein, and Rohani Ganie were responsible for study design and revising. T. Syarfina, and Mahriyuni were responsible for data collection. Muhammad Yusuf and Rahmadsyah Rangkuti drafted the manuscript and revised it. All authors read and approved the final manuscript.

Funding

This work receives no funding.

Competing interests

The authors declare that there is no competing interest.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.