المجلة العراقية للبحوث الإنسانية والإجتماعية والعلمية
Iraqi Journal of Humanitarian, Social and Scientific Research
Print ISSN 2710-0952 Electronic ISSN 2790-1254
العـــدد 17A أيـار 2025
No.17A May 2025

A Corpus Linguistic Study of Lexical Density and Readability of Selected Short Stories

Assistant Professor Dr. Saza Ahmed Fakhry Abdulla
English Department, College of Language, University of Sulaimani
Email: saza.fakhry@univsul.edu.iq
[ORCID iD] https://orcid.org/0000-0001-6228-0982
Assistant Lecturer Shayma Mahmud Ahmed
English Department, College of Language, University of Sulaimani
Email: shayma.ahmed@univsul.edu.iq

Abstract

This paper presents a corpus-based quantitative linguistic analysis of the lexical density (LD) and readability of selected short stories. LD is a key aspect of text readability, and this study posits that the content words in the selected short stories are used to measure the level of LD and readability. Lexical density is a quantitative measure used in corpus linguistics to analyze the complexity of language in texts. Specifically, it investigates the ratio of content words, such as nouns, verbs, adjectives, and adverbs, to the total word count in various text types. This study employed a quantitative research method to manually examine the levels of LD and readability in six selected short stories. For the theoretical framework, Eggins' (2004) equation was used to measure LD, while Khamahani's (2015) framework was applied to evaluate readability. The selected stories include *The Hare and the Tortoise*, *Lazy John*, *The Elephants and the Ants*, *The Dog and the Bone*, *The Thirsty Crow*, and *The Fox and the Grapes*. During data analysis, content words were separated from function words and prepared for the application of the equations. The results revealed that the level of LD in *The Hare and the Tortoise* (50.66%), *Lazy John* (53.3%), and *The Elephants and the Ants* (51.02%) was high, indicating that these stories are lexically dense. Additionally, their readability levels were classified as hard. In contrast, *The Dog and the Bone* (47.61%), *The Thirsty Crow* (45.9%), and *The Fox and the Grapes* (47.36%) had lower LD levels and were classified as easy to read. Among the lexical categories, nouns were the most frequently used, appearing 95 times compared to other content words in the selected short stories.

Keywords: lexical density (LD), readability, lexical category, short story, content words, corpus linguistics

دراسة لغوية للكثافة المعجمية وسلاسة القراءة لعدد من مختارات القصص القصيرة

الاستاذ المساعد دكتور سازه احمد فخري عبدالله
قسم اللغة الانكليزية – كلية اللغات – الجامعة السليمانية
البريد الالكتروني : saza.fakhry@univsul.edu.iq
https://orcid.org/0000-0001-6228-0982: Ordcid

المدرس المساعد شيماء محمود احمد
قسم اللغة الانكليزية – كلية اللغات – الجامعة السليمانية
shayma.ahmed@univsul.edu.iq

الملخص
تقدم هذه الدراسة تحليلاً لغويًا كميًا قائم على مجموعة من النصوص لمعرفة الكثافة المعجمية وقابلية القراءة في مجموعة مختارة من القصص القصيرة . تعد الكثافة المعجمية جانبًا رئيسيًا من قابلية قراءة النص، وتفترض هذه الدراسة أن الكلمات التي تحتوي على المحتوى في القصص القصيرة المختارة تُستخدم لقياس مستوى الكثافة المعجمية وقابلية القراءة. تعتبر الكثافة المعجمية مقياس كمي يستخدم في علم اللغة لتحليل تعقيدات اللغة في النصوص. على وجه التحديد، تبحث في نسبة الكلمات التي تحتوي على المحتوى، مثل الأسماء والأفعال والصفات والظروف، إلى إجمالي عدد الكلمات في أنواع النصوص المختلفة. قامت هذه الدراسة بإستخدام طريقة بحث كمية للتأكد من مستويات الكثافة المعجمية وقابلية القراءة يدويًا في ست قصص قصيرة مختارة. بالنسبة للإطار النظري، تم إستخدام معادلة إيجينز (2004) لقياس الكثافة المعجمية، بينما تم تطبيق إطار خاماهاني (2015) لتقييم قابلية القراءة. تتضمن القصص المختارة الأرنب والسلحفاة، جون الكسول، الفيلة والنمل، الكلب والعظمة، الغراب العطشان، والثعلب والعنب. أثناء تحليل البيانات، تم فصل كلمات المحتوى عن الكلمات الوظيفة وإعدادها لتطبيق المعادلات. أظهرت النتائج بأن مستوى صعوبات التعلم في الأرنب والسلحفاة (50.66٪)، جون الكسول (53.3٪)، والفيلة والنمل (51.02٪) كان مرتفعًا، مما يدل على أن هذه القصص كثيفة المفردات. بالإضافة إلى ذلك، تم تصنيف مستويات قابلية القراءة على أنها صعبة. في المقابل، كان لدى الكلب والعظمة (47.61٪)، والغراب العطشان (45.9٪)، والثعلب والعنب (47.36٪) مستويات صعوبات التعلم أقل وتم تصنيفها على أنها سهلة القراءة. من بين الفئات المعجمية، كانت الأسماء الأكثر استخدامًا، حيث ظهرت 95 مرة مقارنة بكلمات المحتويات الأخرى في القصص القصيرة المختارة.

الكلمات المفتاحية: الكثافة المعجمية، قابلية القراءة، القصة القصيرة، كلمات المحتوى، علم لغة النصوص

1. Introduction

It is undisputed that reading is a primary means of acquiring knowledge from texts (Sholichatun, 2011). According to Gunning (1952), learning is hindered when learners encounter excessive unfamiliar vocabulary during reading, as it impedes overall comprehension. Text readability depends on multiple factors, including lexical density (LD)—the ratio of content words to the total word count in a text. A higher proportion of content words typically indicates greater informational density (Sholichatun, 2011), meaning texts with high LD are generally more complex and challenging to comprehend than those with low LD.

However, despite existing research on LD, no prior studies have measured lexical density in selected children's short stories, leaving a gap in the literature. Therefore, this study aims to (1) assess the readability of children's literature, (2) measure the LD of chosen short stories, and (3) identify the most frequently used content words.

This study seeks to address the following **research questions**:

1. How is LD manifested in the selected short stories?

2. What is the readability level of the selected short stories?
3. Which content words (nouns, verbs, adjectives, and adverbs) are used most frequently compared to others?

The **research gap and problem statement** of this study include that although interest in corpus linguistics has grown in recent years, few studies have explored the relationship between readability and LD in short stories, particularly in children's literature. This study aims to bridge that gap by analyzing how LD and readability influence children's comprehension and engagement with selected short stories. A key issue arises from authors' potential lack of awareness in balancing LD and readability with the target age group's linguistic abilities. If children's stories contain excessive lexical density without appropriate readability adjustments, comprehension may be hindered.

The study tests the following **hypotheses**:

1. A significant relationship exists between LD and readability in the selected short stories.
2. LD directly correlates with readability (i.e., higher LD corresponds to lower readability).
3. Verbs will appear more frequently than other content words (nouns, adjectives, adverbs).

2. Theoretical Framework

Numerous studies have examined **LD** across different linguistic contexts. Johansson (2008) found that written texts generally exhibit higher LD than spoken discourse. Similarly, Fan and Thomas (2013) analyzed LD and readability in English textbooks, revealing that upper-intermediate texts had lower LD than those designed for elementary, pre-intermediate, and intermediate levels.

Further contributing to this field, Nesia and Ginting (2014) conducted a qualitative and descriptive analysis of LD in senior high school English reading materials, specifically examining the *Look Ahead* textbook series. Beyond written texts, LD has also been investigated in spoken discourse. For example, Alami, Sabbah, and Iranmanesh (2013) compared gender differences in LD and concluded that male and female speech patterns displayed nearly identical lexical density, indicating no significant gender-based variation. More recently, Kembaren and Aswani (2022) assessed the LD and readability of five *New York Times* articles using qualitative methods. Their findings suggested that the texts demanded advanced reading comprehension, making them most appropriate for readers aged 18 and above.

In contrast, Bakuuro (2024) applied Ure and Halliday's LD formula to analyze the relationship between LD and text complexity in Ghanaian senior high school English materials. The study revealed an inverse correlation between LD and readability: lower LD corresponded to higher readability, while higher LD was

associated with lower readability. These findings suggest that low readability often linked to high LD may hinder students' academic progress, as text comprehension becomes more challenging with increased LD.

In relation to discourse analysis, Suminih, Hidayani, and Perdana (2024) used critical discourse analysis to investigate lexical density in reading passages from English for Nusantara, a junior high school textbook. A high lexical density of 11.5 was found in their research, suggesting a higher concentration of rich vocabulary and significant content. This promotes a more contextually aware curriculum design by implying that teachers can use these books to help students develop critical thinking skills, cultural awareness, and a stronger bond with their background. Similarly, Ayu and Simatupang (2022) used descriptive qualitative approaches to examine the lexical density in Virginia Woolf's Mrs. Dalloway. According to their analysis, the novel's lexical density was 51%, with nouns making up the largest percentage at 40%. The study found that the novel's narrative complexity is reflected in a variety of lexical items, such as nouns, verbs, adverbs, and adjectives.

The readability and LD of a few children's stories from The Brothers Grimm Fairy Tales: Bath Treasury of Children's Classics were investigated by Faradillah (2024) in a different study. Six stories were examined using content analysis and a qualitative approach. Faradillah used the Flesch Reading Ease (FRE) formula to evaluate readability and Ure's method to quantify lexical density. The results showed that two of the six stories had a high lexical density and four had a low lexical density. Two stories were categorized as "Fairly Easy" (fit for readers in the seventh grade), while four stories were categorized as "Easy" (fit for children in the sixth grade). This suggests that four of the six examined stories are appropriate for children's literature since they exhibit traits of spoken language.

What distinguishes this study from previous research is its focus on selected children’s short stories that have not been previously investigated. Additionally, the study employs Eggins's (2014) and Khamahani’s (2015) analytical research tools, providing a valuable methodological approach.

2.1 Lexical Density (LD)

Ure (1971) was the first to propose the idea of lexical density (LD). Lexical density is a metric used in computational linguistics to quantify the intricacy and organization of human communication in a language (Halliday, 1985). According to Ure (1971) and Eggins (2004), LD is defined as the proportion of lexical words in a text to the total number of words. As Johansson (2009, p. 146) explains, LD is “a term most often used to describe the proportion of content words: nouns, verbs, adjectives, and adverbs (Leech and Short, 2007, pp. 60–61) to the total number of words.” Halliday (1989, p. 80) further notes that written language is typically twice as lexically dense as spoken language. Content words, which include nouns, verbs,

adjectives, and adverbs, do not encompass non-lexical, grammatical, or functional word classes such as conjunctions, articles, prepositions, and non-lexical verbs and adverbs (Halliday, 1985b). In other words, content words belong to open classes, unlike function words, which are part of closed classes (e.g., Cruse, 2011). However, some linguists define LD as the ratio of open-class words to closed-class words. For instance, Marlia et al. (2021) define lexical density as "a statistic that compares the quantity of content and function terms in a document." It is important to note that the current study aligns with Halliday's (1985b) definition, which measures LD by comparing the number of content words to the total number of words, rather than to the number of functional words. This approach is supported by evidence suggesting that texts with low lexical density are generally easier to comprehend than those with high lexical density (Stubbs, 2004). To calculate LD, it is essential to distinguish between lexical and grammatical categories. Grammatical elements, also known as function words, include determiners (e.g., articles, pronouns), prepositions, conjunctions, certain classes of adverbs, and finite verbs. A closed system consists of these components (Halliday & Matthiessen, 2014).

On the other hand, words that have semantic content and add to the meaning of the phrase in which they appear are known as lexical items, or content words. Radford (2004) states that nouns, verbs, adjectives, adverbs, and prepositions are examples of lexical categories. However, prepositions are excluded from this list in the current study because they lack semantic substance when used independently. Therefore, nouns, verbs, adjectives, and adverbs are the lexical items used as the numerator in the LD formula for this work. Lexical density is a linguistic measure that quantifies the ratio of content words (lexical items), such as nouns, verbs, adjectives, and adverbs, to the total number of words in a text. It provides a numerical representation of the proportion of meaningful words relative to other linguistic components (Suminih, Hidayani, and Perdana, 2024). Two primary formulas are used to calculate LD: Ure's (1971) formula, which is LD = (Number of lexical items / Total number of words) × 100, and Halliday's (1985) formula, which is LD = (Number of lexical items / Total number of clauses) × 100.

2.2 Lexical Density & Readability

Rizkiani, Mahdi, and Sujatna (2022) state that in addition to LD, a text's readability must be measured to help teachers create lesson plans that follow the curriculum. Pikulski (2002) argues that readability is "the level of ease or difficulty of a text material that can be understood by certain readers and reads the text for a specific purpose." Moreover, LD and text readability are inversely correlated; the denser a text, the more difficult it is to comprehend and process (Harrison & Bakker, 1998; Stubbs, 2004; Castello, 2008). According to Sujatna, Yuyu, and Rahmat (2017), to

improve readability, children's short stories should have low LD. On the words of Richard, Richards, Platt, J., & Platt, H. (1992), readability "depends on several factors including the average length of sentences, the number of new words contained, and the grammatical complexity of the language used in a passage." To Dubay (2006), readability is the quality that makes certain writings simpler to read and understand than others. To Sholichatun (2011), Texts with a lexical density score of 40–50% are considered low density, whereas those with a score of 50–60% are considered highly dense. The degree of difficulty of a text can be ascertained by analyzing its readability. A text's lexical density affects the reader's understanding and difficulties (Amer, & Baarah, 2021; Rizkiani et al., 2022; Aswani et al., 2023). A quantitative research method is used with the equation proposed by Eggins (2004) to measure the level of LD in the selected short stories. This can be shown as follows:

Formula:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\%$$ (Eggins, 2004)

2.3 Corpus Linguistics and Lexical Density

McEnery and Hardie (2012) discuss the balance between manual and computational methods in corpus linguistics. They argue that while computational tools are essential for handling large datasets, manual analysis is still valuable for interpreting results and providing deeper insights.

Lexical density differences between spoken and written language, formal and informal contexts, or distinct languages can be revealed with the aid of corpus linguistics. Corpus linguistics is the systematic study of language using large, structured text collections called corpora (plural: corpus). These corpora are analyzed to identify trends, patterns, and frequencies in language use (Bennett, 2010). In relation to lexical density, corpus-based analytical approaches are rather useful in generating empirical and quantitative stylistic descriptions of literary works (Ina'am Abdul-Kadhim and Hussein, 2020). Corpus linguistics is used to study grammar, vocabulary, discourse, and even lexical density across different texts or genres. Lexical density is a means of calculating the proportion of lexical words in a text or corpus, by expressing it as a percentage (Ure 1971). Due to the numerous debates surrounding the statistical distinctions between lexical items and function words, the computations of these terms are essentially complex. In light of this, corpus linguistics provides a collection of statistical techniques that address figuring out how many of those words and items there are. To determine the lexical density of a given text, these figures must then be entered into a statistically sound yet practical formula. The final phase in this procedure will be the identification of lexical density (Ina'am Abdul-Kadhim and Hussein, 2020).

3. Methodology

A quantitative research method is used with the equation proposed by Eggins (2004) to measure the level of LD in the selected short stories. This can be shown as follows:

Formula:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\%$$ (Eggins, 2004)

Regarding this measurement, the level and the obtained score of LD would be checked against 50%. If the score of LD exceeds 50%, it indicates that the text is lexically dense (Khamahani, 2015). The **data collection** is sourced from *10 Lines of Short Stories with Moral Lessons for Kids*. This collection consists of 10 short stories, out of which six have been selected for data analysis. The selected texts used to investigate LD and readability, focusing on content words such as nouns, verbs, adjectives, and adverbs. The stories are *The Hare and the Tortoise, Lazy John, The Fox and the Grapes, The Elephants and the Ants, The Dog, and the Bone*, and *The Thirsty Crow*. Manual Counting is used for the measurement as the corpus was small and also the selected texts. In the first place, the total words were calculated using Microsoft Word and the content words were then counted manually to obtain the number of each content word in each text. Meanwhile, the ratio of the LD is also taken from the equation proposed by Eggins (2004). The data analysis relies on Eggins (2004) to investigate the LD of the selected six short stories and Khamahani (2015) to measure the readability level of those stories.

4. Discussion and Analysis

The selected short stories will be examined below to show the LD and the readability level for the content words: nouns, verbs, adjectives, and adverbs.

1. The Hare and the Tortoise

There was once a hare who was friends with a tortoise. One day, he challenged the tortoise to a race. Seeing how slow the tortoise was going, the hare thought he'd win this easily. So, he took a nap while the tortoise kept on going. When the hare woke, he saw that the tortoise was already at the finish line. Much to his chagrin, the tortoise won the race while he was busy sleeping.

Table 1 illustrates the content words (content words) in Text 1

Short Text 1				
Content words				
Nouns	**Verbs**	**Adjectives**	**Adverbs**	**Total**
16	14	3	5	38

The total number of words used in the above short text is 75. The total number of content words captured in the text is 38, consisting of 16 nouns, 14 verbs, 3 adjectives, and 5 adverbs. Considering the ratio of the total number of content words to that of words in the text, the result for the LD of text 1 is 50.66%. This result is calculated using the following equation:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\% \quad \text{(Eggins, 2004)}$$

$$Lexical\ Density = \frac{38}{75} x\ 100\% = 50.66\%$$

As shown in the equation above, the LD of text 1 is 50.66% which is slightly above 50%. This result indicates that the text is lexically dense.

2. The Dog and the Bone

Once, there was a dog who wandered the streets night and day in search of food. One day, he found a big juicy bone, and he immediately grabbed it in his mouth and took it home. On his way home, he crossed a river and saw another dog with a bone in its mouth. He wanted that bone for himself, too. But as he opened his mouth, the bone he was biting fell into the river and sank. That night, he went home hungry.

Table 2 illustrates the Content words in Text 2

Short Text 2				
Content words				
Nouns	**Verbs**	**Adjectives**	**Adverbs**	**Total**
21	13	3	3	40

The total number of words used in the above short text under study is 84. The total number of content words captured in the text is 40, consisting of 21 nouns, 13 verbs, 3 adjectives, and 3 adverbs. Considering the ratio of the total number of content words to the total number of words in the text, the lexical density (LD) of Text 2 is 47.61%. This result is calculated using the following equation:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\% \quad \text{(Eggins, 2004)}$$

$$Lexical\ Density = \frac{40}{84} x\ 100\% = 47.61\%$$

As shown in the equation above, the LD of text 2 is 47.61%, which is less than 50% and is considered a low level of LD.

3. The Thirsty Crow

After flying a long distance, a thirsty crow wandered the forest searching for water. Finally, he saw a pot half-filled with water. He tried to drink from it, but his beak wasn't long enough to reach the water inside. He then saw pebbles on the ground, and one by one, he put them in the pot until the water rose to the brim. The crow then hastily drank from it and quenched his thirst.

Table 3 illustrates the Content words in Text 3

Short Text 3				
Content words				
Nouns	**Verbs**	**Adjectives**	**Adverbs**	**Total**
15	14	3	2	34

The total number of words used in the above text understudy is 74. The total number of lexical phrases (items) captured in the text is 34, consisting of 15 nouns, 14 verbs, 3 adjectives, and 2 adverbs. Considering the ratio of the total number of content words to that of words in the text, the result for the LD of text 3 is 45.9%. This result is calculated using the following equation:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\% \quad \text{(Eggins, 2004)}$$

$$Lexical\ Density = \frac{34}{74} x\ 100\% = 45.9\%$$

As shown in the equation above, the lexical density (LD) of Text 3 is 45.9%, which is below 50%. This indicates a low level of lexical density.

4. Lazy John

There was a boy named John who was so lazy he couldn't even change his clothes. One day, he saw the apple tree in their yard was full of fruits. He wanted to eat some apples, but he was too lazy to climb the tree and take the fruits. So he lay down underneath the tree and waited for the fruits to fall off. John waited until he was starving, but the apples never fell.

Table 4 illustrates the Content words in Text 4

Short Text 2				
Content words				
Nouns	**Verbs**	**Adjectives**	**Adverbs**	**Total**
15	17	3	5	40

The total number of words used in the above text being studied is 75. The total number of lexical phrases (items) captured in the text is 40, consisting of 15 nouns, 17 verbs, 3 adjectives, and 5 adverbs. Considering the ratio of the total number of content words to that of words in the text, the result for the LD of text 4 is 53.3%. This result is calculated using the following equation:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\% \quad \text{(Eggins, 2004)}$$

$$Lexical\ Density = \frac{40}{75}\ x\ 100\% = 53.3\%$$

As shown in the equation above, the lexical density (LD) of Text 4 is 53.3%, which is above 50%, indicating that it is a lexically dense text.

5. The Fox and the Grapes

Once, there was a hungry fox who stumbled upon a vineyard. After seeing the round, juicy grapes hanging in a bunch, the fox drooled. But no matter how high he jumped, he couldn't reach for it. So he told himself that it was probably sour and left. That night, he had to sleep on an empty stomach.

Table 5 illustrates the Content words in Text 5

Short Text 5				
Content words				
Nouns	**Verbs**	**Adjectives**	**Adverbs**	**Total**
8	11	5	3	27

The total number of words used in the above text understudy is 57. The total number of lexical phrases (items) captured in the text is 27, consisting of 8 nouns, 11 verbs, 5 adjectives, and 3 adverbs. Considering the ratio of the total number of content words to that of words in the text, the result for the LD of text 5 is 47.36%. This result is calculated using the following equation:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\% \quad \text{(Eggins, 2004)}$$

$$Lexical\ Density = \frac{27}{57}\ x\ 100\% = 47.36\%$$

As shown in the equation above, the LD of Text 5 is 47.36% which is slightly below 50% which means it has a low level of lexical density.

6. The Elephant and the Ants

There was once a proud elephant who constantly bullied smaller animals. He would go to the anthill near his home and spray water at the ants. The ants, with their size, could do nothing but cry. The elephant just laughed and threatened the ants that he would crush them to death. One day, the ants had enough and decided

to teach the elephant a lesson. They went straight into the elephant's trunk and started biting him. The elephant could only howl in pain. He realized his mistake and apologized to the ants and all the animals he bullied.

Table 6 illustrates the Content words in Text 6

Short Text 6				
Content words				
Nouns	**Verbs**	**Adjectives**	**Adverbs**	**Total**
22	19	4	5	50

The total number of words used in the above text is 98. The total number of lexical phrases (items) captured in the text is 50, consisting of 22 nouns, 19 verbs, 4 adjectives, and 5 adverbs. Considering the ratio of the total number of content words to that of words in the text, the result for the LD of text 6 is 51.02%. This result is calculated using the following equation:

$$Lexical\ density = \frac{Number\ of\ Lexical\ items}{Total\ number\ of\ words}\ x\ 100\% \quad \text{(Eggins, 2004)}$$

$$Lexical\ Density = \frac{50}{98}x\ 100\% = 51.02\%$$

As shown in the equation above, the LD of Text 6 is 51.02%, which is slightly above 50%. Therefore, it is considered a lexically dense text."

4. Findings and Conclusion

The findings of the analysis can be shown in the table below:

No.	Selected Short Stories	Ratio of Content words	Level of Lexical Density	Level of Readability
1.	The Hare and the Tortoise	50.66%	High	Hard
2.	The Dog and the Bone	47.61%	Low	Easy
3.	The Thirsty Crow	45.9%	Low	Easy
4.	Lazy John	53.3%	High	Hard
5.	The Fox and the Grapes	47.36%	Low	Easy
6.	The Elephants and the Ants	51.02%	High	Hard

The level of LD and ration in *The Hare and the Tortoise* (50.66%), *Lazy John* (53.3%). *The Elephants and the Ants* (51.02%) were high and their levels of

readability were hard and, in *The Dog and the Bone* (47.61%), *The Thirsty Crow* (45.9%), and *The Fox and the Grapes* (47.36%), were low and easy. From the table above, it is revealed that 3 of the short stories have a high level of LD. *The Hare and the Tortoise, Lazy John,* and *The Elephants and the Ants* are lexically dense, as well as their levels of readability were hard while 3 of them have a low level of LD, namely *The Dog, and the Bone*, *The Thirsty Crow*, *and The Fox and the Grapes*. As quoted earlier, LD was inversely proportional to text readability; the denser a text is, the harder it is to process and understand. Therefore, the above 3 stories have low levels of LD and are easy to read while the above 3 stories that are lexically dense are hard in the level of readability.

Beginning with answering the aforementioned research questions, the first query was to figure out the level of LD for the 6 selected short stories, the level of LD is low for 3. However, the other 3 are lexically dense. Concerning the second query, to find out the level of readability, as seen in the table provided above, 3 of the stories are low in the level of readability while the other 3 of have a high level. The third query was which lexical category is found the most. The answer is the 'nouns' which are found 95 times compared to the other content words that are seen in the table below:

No.	Selected Short Stories	Nouns	Verbs	Adjectives	Adverbs
1.	The Hare and the Tortoise	14	16	3	5
2.	The Dog and the Bone	21	13	3	3
3.	The Thirsty Crow	15	14	3	2
4.	Lazy John	15	17	3	5
5.	The Fox and the Grapes	8	11	5	3
6.	The Elephants and the Ants	22	19	4	5
7.	Total	95	90	21	23

Based on the research questions presented in Section 1, it was found that the readability of some of the short stories is not appropriate for children, as the high level of lexical density (LD) makes the texts difficult to read. In conclusion, this study explored the examination of lexical density and readability in six short stories. As presented above, three of the short stories have a low level of readability, while the other three have a high level. Since lexical density and readability are inversely proportional, the three stories with low readability had a high level of LD, whereas the other three were lexically dense. Therefore, the results indicate that half of the short stories are hard to read for children, while the other half are easy. Finally, the last objective of this study was to identify the most frequent content words among the six short stories. Nouns, which appeared 95

times, were found to be the most frequent content words compared to the other categories (verbs, adjectives, and adverbs).

Acknowledgment

Our profound appreciation goes out to everyone who helped us finish this study. First, we want to express our gratitude to the journal editors and reviewers for their insightful comments and helpful recommendations, which have greatly improved the quality of this work. We are especially grateful to the Department of English Language for creating a setting that supports scholarly inquiry. Furthermore, we would like to thank those scholars and researchers whose work has been beneficial and who have made significant contributions to the fields of lexical density and readability.

Data Availability Statement: The manuscript includes all the data used in the study.

Conflict of Interest Statement: (The authors confirm that there are no conflicts of interest that could affect the content of this research.

Funding Statement: This research is fully funded by the authors without any financial support from other entities.

References

Alami, Manizheh., Sabbah, Maryam., & Iranmanesh, Muhammad. 2012. MaleFemale Discourse Difference in Terms of Lexical Density. Research Journal of Applied Sciences, Engineering and Technology, 5(23): 5365- 5369.

Ayu, S.N. and Simatupang, E.C., 2022. Lexical Density in The Novel "Mrs. Dalloway" by Virgina Woolf. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, *5*(3), pp.18845-18852.

Bennett, G.R., 2010. Using CORPORA in the Language Learning Classroom–corpus linguistics for teachers. University of Michigan.

Bakuuro, J.2024. In the Belly of Text Complexity: Unravelling the Nexus between Lexical Density and Readability, Athens Journal of Philology.Vol. X, No. Y.

Amer, M. A. B., & Baarah, H. A. (2021). Readability and Lexical Density of Reading Sections of Tenth Grade English Textbooks in Jordan and Sultanate of Oman: A comparative Study. Budapest International Research and Critics in Linguistics and Education (BirLE) Journal, 4(3), 1138-1148.

Castello, E. (2008). Text complexity and Reading Comprehension Tests. Linguistic Insights: Studies in Language and Communication, No. 85. Bern: Peter Lang.

Dubay, W.H. (2006). Unlocking language: The classic readability studies.
California: Impact information.

Eggins, S. (2004). An Introduction to Systemic Functional Linguistics (2nd ed.). London: Continuum.
Faradillah, **S.** (2024). *Lexical Density and Readability of Selected Children Stories*. Repositori Institusi USU, Universitas Sumatera Utara. Retrieved from https://repositori.usu. ac.id/ handle /123456789/93549.
Gunning R (1952) The Technique of Clear Writing. New York: McGraw-Hill.
Halliday M.A.K. (1985b) Spoken and Written Language. Waurn Ponds, Vic: Deakin University.
Halliday, M. A. K. (1989). Spoken and Written Language (2nd ed.). Oxford: Oxford University
Press.
Halliday, M. A. K. & Matthiessen, C.M.I.M. (2014). Halliday's Introduction to Functional
Grammar (4thed.). London: Routledge
Harrison, S. & Bakker, P. (1998). Two New Readability Predictors for the Professional Writer: Pilot Trials. Journal of Research in Reading, 21(2), 121–138.
Ina'am Abdul-Kadhim and Hussein, K.S. (2020). A Corpus-Based Stylistic Identification of Lexical Density Profile of Three Novels by Ernest Hemingway." The Sun also Rises"," A Farewell to Arms" and" A Moveable Feast". GRIN Verlag.
Johansson, V. (2008). Lexical diversity and lexical density in speech and writing. Lund University:Lund University Press.
Johansson, V. (2009). Developmental Aspects of Text Production in Writing and Speech. Travaux de l'Institut de Linguistique de Lund, 48. Lund: Lund University
Kembaren, F.R., & Aswani, A.N., (2022). Exploring Lexical Density in the New York Times. English Language, Literature, and Teaching, 7(2), 109-119 doi:10.32528/elite.v7i2.8795
Khamahani G (2015) A Corpus-Based Analysis of Tehran Times and Azeri News Headlines: Focus on Lexical Density and Readability. International Journal of Humanities, Social Sciences and Education 2(1): 12–16.
Leech, G., & Short, M. H. (2007). *Style in Fiction: A Linguistic Introduction to English Fictional* Prose (2nd ed.). London: Longman.
Lukmana, I., & Gunawan, W. (2021, November). Lexical density in UU Cipta Kerja: A case study of controversial articles. In Proceedings of the Tenth International Conference on Languages and Arts (ICLA 2021) (pp. 114-119). Atlantis Press.
McEnery, T. and Hardie, A. (2012). Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press.
Nesia BH, & Ginting SA (2014) Lexical Density of English Reading Text for Senior High School. Medan, Indonesia: University of Medan.

Pikulski.(2002).Readability.Retrievedfromhttps://www.eduplace.com/state/author/pikulski.pdf

Radford , A.2004. *English Syntax, First*. Cambridge: Cambridge University Press.

Richards, J. C., Platt, J., & Platt, H. (1992). Longman dictionary of language teaching and applied linguistics. London:Longman

Sholichatun, S (2011) Content analysis of Reading Materials in 'English on Sky' Textbook for Junior High School. Undergraduate Thesis. Walisongo Institutional Repository.

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.

Suminih, S., Hidayani, S. and Perdana, P.R., 2024. Lexical Density in English Textbook of Junior High School. *Loquen: English Studies Journal*, *17*(2), pp.1-18.

Stubbs, M. (2004). Language Corpora. In A. Davies & C. Elder (Eds.), The Handbook of Applied Linguistics (pp. 106–132). Malden: Blackwell.

Sujatna,E. Yuyu,R. and Rahmat,S.(2017). The Readability Test of the English Children Short Stories. International Conference on Language, Education, Humanities and Innovation 21St & 22nd January, 2017

Štajner, S. & Mitkov, R., 2012, May. Using comparable corpora to track diachronic and synchronic changes in lexical density and lexical richness. In The 5th workshop on building and using comparable corpora (Vol. 17, No. 17, p. 88).

Rizkiani, D., Mahdi, S., & Sujatna, E. T. S. (2022). Lexical Density and Readability of the Facil's 'Advance Learning Textbook' for Indonesian High School Students. Al-Ishlah: Journal

Ure, J. (1971) 'Lexical density and register differentiation', in G. Perren and J. L. M. Trim (eds), Applications of linguistics, London:Cambridge University Press, pp. 443–52.