

Exploring Lexical Density in the New York Times

*Farida Repelita Kembaren¹, Ade Novira Aswani²

¹²Universitas Islam Negeri Sumatra Utara, Indonesia
(*my.paper.usu@gmail.com)

First Received: 17-11-2022

Final Proof Received: 29-11-2022

Abstract

The goal of the study was to evaluate the lexical density and readability of five news articles from The New York Times. A qualitative content analysis methodology was employed in conducting the research. Methods from Ure (1971) and Halliday (1985) were used to analyze the lexical density, and Flesch's (1948) Reading Ease Score and Grade Level were used to evaluate the readability of the news articles. The findings showed that The New York Times news articles had lexical densities that were higher than 50%, significantly lower readability scores than 50, and grade levels greater than 12. The percentages of lexical density were higher than Ure's (1971) minimum score for written texts. The readability score also revealed that the news articles required a high level of reading comprehension and were best understood by readers who were at least 18 years old or at least in college. The study came to the conclusion that The New York Times is one of the most trustworthy and reputable news sources.

Keywords: Lexical density; readability; grade level; newspaper; The New York Times.

Newspapers are crucial to most people's lives. It is one of the reasonably priced forms of communication in which individuals of all socioeconomic classes can reach out each morning to receive the daily paper and learn about the day's happenings across the world. In our digital age, people can use their gadgets to obtain information from all around the world. Because the news media has an impact on everyday life in modern society, they have the ability to aid in the analysis of empirical evidence of complex human-nature relationships (Biagi, 2012; Jaung & Carrasco, 2022). The media have a critical role in distributing knowledge regarding human-nature relationships, calling public attention to the issue, shaping public

debate, supporting changes in consumptive behavior, integrating research and policy in the public domain, and influencing policies and regulations (Boykoff & Yulsman, 2013). For information, people rely on both printed and internet outlets. We can use The New York Times as an example of an online newspaper in this case.

Based on the data from a Reuters Institute survey in 2020, as many as 20% of newspaper readers in the United States paid for online access to their favorite publications. In 2020, newspapers were used by 16% of adults in the United States to obtain information. In Asia Pacific, Indonesia was the second largest consumer of paid news in 2021. According to

the findings of an investigation conducted by the Reuters Institute, Hong Kong had the highest percentage of people who paid for online news among Pacific countries, with 23% of respondents. With 19% of respondents admitting to paying for premium online news material, Indonesia came in second. Readers could obtain excellent and comprehensive news by paying for premium news.

Reading is an essential skill for everyone in the world today. If someone wants to read articles or books without having to wait for long periods of time for translations in their language, they must be able to read in English because many articles, books, manuscripts, and journals are published in English now. It will take a long time to translate them, and once done, the information will no longer be new. The New York Times is one of the websites that publishes news in English. Reading comprehension is one of the struggles for every reader, even those who read The New York Times as an online newspaper.

According to digital newspaper readership data, The New York Times was ranked as the first of the top 200 newspapers in the world by the 2019 Newspaper Web Rankings. The New York Times was also chosen by 39% of those who signed up for digital publications in 2020. This is one of the reasons why the researchers chose The New York Times news articles to be the research object.

According to Deepublish Publisher, there are several characteristics of good news texts, including: (1) actual; (2) factual; (3) objective; (4) the data shown is complete; and (5) the sentences are short, concise, and clear. The points (4) and (5) are related to lexical density. The news text must show complete data in short, concise, and clear sentences because every newspaper has limited words. That is why the sentences must have a high lexical density.

The lexical density reveals the intricacy of written text. Density refers to the amount of information conveyed through

written language. The percentage of lexical elements or content words in the overall text is known as lexical density (Halliday, 1985). It shows the connection between lexical and grammatical elements. Nouns, verbs, adverbs, and adjectives are considered lexical or content words. Meanwhile, prepositions, finite verbs, classes of adverbs, pronouns, conjunctions, and determiners are considered grammatical words or glue words. Furthermore, Ure (1971) believes that the lexical density of the vast majority of spoken texts is less than 40%, but the lexical density of the vast majority of written texts is 40% or higher. It means that lexical density is the depiction of content words in a text or sentence, and that a text with a high percentage of lexical density is a written text (Hanafiah & Yusuf, 2016). Sticky sentences, on the other hand, result from the use of a high percentage of grammatical terms. Sentences with a high percentage of grammatical words are referred to as sticky sentences (Qub'a, Guba, & Hasan, 2022). The use of several of these grammatical words will require readers to go through more text to get the genuine meaning because they are considered to be empty spaces in writing.

The lexical density of a text reduces when it has fewer lexical items than grammatical ones. On the other hand, a text has a high lexical density if it has more lexical items than grammatical ones, which makes it more challenging to understand yet still contains a wealth of information (Rizkiani, Mahdi, & Sujatna, 2022). The key factor that affects how difficult or simple it is for a reader to understand a text is lexical density (Amer, 2021). The number of lexical items in a text can influence its length and level of reading difficulty. It becomes harder to read a document when more lexical items are included in it. Reading a passage gets challenging as the percentage of lexical density increases significantly. Naturally, the complexity will also be influenced by the specific lexical items employed and by how they are placed within the grammatical framework (Halliday and Martin, 1993). Lexical density

has an impact on the reader's comprehension of the content. The readers' comprehension of how the lexical density in the texts they are reading would affect their capacity to learn English is influenced by certain new terms. According to Gerot and Wignel (1994), lexical density is a gauge of how much content information is contained in a clause (or in a text).

Lexical density is quite related to the readability of a text. One of the most crucial factors to take into account while choosing a suitable piece for readers is readability. The term "readability" is defined differently by several experts. In reality, the readability of a piece of writing is influenced by many more factors than just the length of the sections or the words themselves, such as how engaging and appealing the writing is. The readability of texts was also examined to gauge the degree of literary complexity and determine the reading proficiency needed to understand the text (Woods, Moscardo, & Greenwood, 1998). The average sentence length and the average number of syllables per word are used in calculations to determine readability (Qub'a, Guba, & Hasan, 2022).

Analyzing the text readability is one method for assessing the text suitability. According to Richard (1992), readability refers to how easily a material may be read and understood. The comprehension and reading of a text for the reader determines how readable it is (Djuharie, 2008). By evaluating the readability of the reading materials, we may locate a reading selection that corresponds to the readers' grade level (Anderson & Anderson, 1998). The Readability of a text is one of its most well-known aspects. Readability was described as the ease of understanding or comprehension owing to the style of writing (DuBay, 2004). According to DuBay (2007), readability is also the ease of reading provided by the choice of material, style, design, and organization that fits the prior knowledge, reading skill, interest, and motivation of the audience. Ruddell (2008) stated in accordance

with those criteria that readability relates to the difficulty level of text materials, an area of constant concern for instructors, students, textbook authors, and publishers.

Readability is the extent to which texts are regarded to be readable by a target demographic. Certain algorithms that count the average sentence length and the average number of syllable per word are used to determine readability. Flesch (1948) devised Flesch Reading Ease Score. This is a scale from 0 to 100 that indicates how difficult a text is to read. The text becomes more difficult to read as the number decreases. These are the grades in the United States of America. As it can be seen in **Table 1** the Flesch Reading Ease Score Table.

Table 1. Flesch Reading Ease Score

Score	School Level	Ease of Readability
90-100	5th grade	Very easy to read
80-90	6th grade	Easy to read
70-80	7th grade	Fairly easy to read
60-70	8th & 9th grade	Plain English
50-60	10th & 11th grade	Fairly difficult to read
30-50	College	Difficult to read
10-30	College graduate	Very difficult to read
0-10	Professional	Extremely difficult to read

Flesch devised another formula to arrive at an exact index of text readability. This formula and the original Flesch Reading Ease Formula correlated to each other. The Reading Ease formula was modified to produce a grade-level score in 1976 as part of a study commissioned by the U.S. Navy. This well-known equation is also referred to as the Flesch-Kincaid Grade Level, Flesch Grade Scale formula, or Kincaid formula. The following equation yields the Flesch-Kincaid Grade Level.

Readability Formula:

$$RE = 206.835 - (1.015 \times ASL - 84.6 \times ASW)$$

RE : Readability Ease
 ASL : Average Sentence Length
 ASW : Average number of Syllables per Word

The commonly used readability formula Flesch-Kincaid Grade Level determines the estimated reading grade level of the texts. The US Navy created the formula while collaborating with the previously known formula, Flesch Reading Ease Score. In the past, a table had to be used to convert the formula to the reading grade level. In order to make it simpler to use, an updated version was created in the 1970s. For its technical instruction manuals, the Navy employed it. The formula now has a considerably larger range of uses. For example, in order to understand a work with Flesch-Kincaid level of 11, the readers must be at least a 11th grade students or 17 years old.

In order to conduct this study, the researchers used online text analyzers to find the lexical density and readability of The New York Times news articles, using three methods of analysis proposed by Ure (1971), Halliday (1985), and Flesch (1948). To analyze the lexical density and readability of the news articles, the researchers used two websites, namely Analyze My Writing (<https://www.analyzemywriting.com/>) and the Automatic Readability Checker feature from Readability Formulas (<https://readabilityformulas.com/free-readability-formula-tests.php>). And to check the lexical item categories, the researchers used the online Merriam-Webster Dictionary (<https://www.merriam-webster.com/>).

Analyze My Writing was an online text content analyzer. The researchers used this website to analyze the lexical density of the five article samples. The website was very helpful for the researchers to find the percentage of all parts of speech in the news texts. After analyzing the lexical density, the researchers then accessed the *Readability Formulas*. The website had many features, and one of them was the Automatic Readability Checker, which the researchers used to analyze the

readability and grade level according to Flesch Reading Ease Score and Flesch-Kincaid Grade Level. The website showed the percentage of monosyllabic, disyllabic, three-syllabic, and more syllabic words in the text samples. The researchers then made the realization by pointing out each lexical item used in the sentences, whether the lexical item fell into the noun, adjective, verb, or adverb category. The researchers used the online *Merriam-Webster Dictionary*, which was an English-to-English dictionary that also showed the part of speech of the word we were searching for.

For his study on lexical density, Khamahani (2015) analyzed 200 headlines from the Tehran Times and Azeri News. The results of a comprehensive investigation of two sets of lexical density in two sets of newspaper headlines found that the Azeri News headlines had a higher lexical density index than the Tehran Times headlines.

Madinah (2018) examined lexical metaphors in The Jakarta Post's headlines. The goals of the study were to identify the different varieties of lexical metaphor, to explain how The Jakarta Post uses lexical metaphor in its headlines, and to justify why The Jakarta Post uses lexical metaphor in its headlines.

Wan (2020) compared 200 online news headlines from the New York Times (American English) and China Daily (Chinese English) news websites, with the goal of determining the linguistic complexity of the two sets of data at the lexical and structural levels. More information needs to be gathered.

Din and Ghani (2020) did a corpus-based study to discover the verb patterns that have appeared in Pakistani newspaper headlines. They generated a corpus of 3135 newspaper headlines with a total word count of 28646 words from three online Pakistani English newspapers, including The News, The Nation, and The Dawn, in order to detect different verb patterns utilized in newspaper headlines.

Yang, Zhang, and Yuan (2021) investigated how China Daily (CD), Cable

News (CNN), and the Daily Mail (DM) reported on the COVID-19 pandemic. The mainstream media is a trustworthy source of information regarding COVID-19. The word frequency and lexical diversity statistics revealed how early pandemic reports changed. Sentiment analysis was used to examine the sentiment of all headlines and the sentiment of the individual words in the headlines.

All of the studies listed above have one thing in common: they all analyzed the headlines of various online publications. Although several studies have examined the lexical density of news headlines from online newspapers, only a few studies have analyzed the lexical density of the news content from online newspapers. The goal of this research itself is to analyze the lexical density of five educational articles from The New York Times to determine their readability. This distinguishes this study from others in terms of the research object. It is hoped that this study will help readers choose reading materials that fit all five of the characteristics of good news text that have been mentioned earlier, based on the percentage of lexical density and readability in news articles.

The lexical density and readability of the news texts are the two key aspects on which this study focuses. The first objective is to determine the realization of the lexical density percentage represented in The New York Times articles. The second objective is to determine the realization of the readability represented in The New York Times articles.

Method

This study used qualitative methodology. There were several methods for analyzing qualitative data, and a qualitative content analysis method was used in this research. In this research, the data needed were the number of lexical items or content words and clauses to analyze the lexical density, the sentence length and number of syllables to determine the readability. Five educational articles from The New York Times were the

secondary data source for this study. The data were taken from news articles published in February 2022.

The five educational news articles from The New York Times were examined using two websites, namely Analyze My Writing and Readability Formulas. Analyze My Writing (<http://www.analyzemywriting.com>) was used to analyze the lexical density, meanwhile the Readability Formulas website (<https://readabilityformulas.com/>) was used to analyze the readability and grade level. The data were measured based on the formula by Ure (1971) and Halliday (1985) to find the lexical density of the five articles. After that, to determine the readability of the news articles from The New York Times, the researchers used Flesch's (1948) Reading Ease Formula and Flesch-Kincaid Grade Level. And the last, to make the lexical density realization, the researchers needed to know the lexical items used in each sentence were in what part of speech category, so the researchers also used online Merriam-Webster Dictionary (<https://www.merriam-webster.com/>) to check the part of speech for each word in the sentence examples.

Results and Discussion

As mentioned in the method, the samples for this study were five educational articles from The New York Times published during February 2022. The titles, authors, publication dates of the five articles were shown in **Table 2**.

Table 2. The titles, authors, and publication dates of the sample articles

No	Title	Author	Publication
1.	San Francisco Ousts School Board Members	Amelia Nierenberg	Feb. 16, 2022
2.	You Just Can't Tell the Truth about America Anymore	Jamelle Bouie	Feb. 18, 2022
3.	The Battle over Race, Tradition and an Elite Private School's Mascot	Corey Kilgannon	Feb. 19, 2022

4.	The Lessons of Brooklyn Tech	Bret Stephens	Feb. 1, 2022
5.	Polio, Chicken-pox, Measles ... Covid-19? The Case for Vaccine Mandates in Schools.	Elisabeth Rosenthal	Feb. 7, 2022

In this part, the researchers provided the characteristics of five sample news articles from The New York Times in terms of word count, sentence count, syllable count, lexical density, and readability.

Lexical Density Analysis

This section responded to the first research question: How is the realization of lexical density percentage represented in The New York Times articles?

The parts of speech percentage contained in the entire articles, as well as the results of the lexical density analysis, were shown by the **Table 3**.

Table 3. Lexical density analysis

Part of Speech	Ar 1 (%)	Ar 2 (%)	Ar 3 (%)	Ar 4 (%)	Ar 5 (%)
Nouns	34.66	32.58	33.31	32.19	28.98
Adjectives	6.42	8.61	8.86	8.25	9
Verbs	12.97	12.09	12.7	8.37	11.17
Adverbs	4.36	3.69	2.51	5.9	5.87
Prep.	12.58	15.57	13.37	16.27	14.3
Pronouns	5.01	2.46	2.44	4.13	3.69
A.V.	5.65	2.97	4.51	5.19	5.68
L.D.	58.41	56.97	57.39	54.72	55.02

Ar : Article
A.V. : Auxiliary Verbs
L.D. : Lexical Density

Table 3 shows, the lexical density of the five text samples were around 54% to 58%, with an average of 56.5%. This percentage was much higher than Ure's (1971) minimum percentage of lexical density in written texts. The table also showed that the part of speech used the most in the article samples was noun, which was around 28% to 34%, with an average of 32.34%, that made the news articles in The

New York Times had high percentage of lexical density.

The followings were the examples of sentences using a high percentage of content words in the news text samples.

- Criticism of the board grew stronger** when voters discovered tweets written by the board's vice president, Alison Collins, one of the three ousted members. [Article 1, lexical density = 70.83%, the lexical items used in this sentence were nouns (criticism, board, voters, tweets, board's vice president, Alison Collins, members), verbs (grew, discovered, written, ousted), and adjectives (stronger, one, three)].
- University officials "are often thinking about protecting the school, not protecting the client," said Arthur Caplan, a professor of medical ethics at New York University's Grossman School of Medicine.** [Article 1, lexical density = 72.41%, the lexical items used in this sentence were nouns (university officials, school, client, Arthur Caplan, professor, ethics, New York University's Grossman School, Medicine), adverbs (often, not), verbs (thinking, protecting, said), and adjective (medical)].
- It **abolished requirements based primarily on grades and test scores, instead implementing a lottery system.** [Article 1, lexical density = 73.33%, the lexical items used in this sentence were verbs (abolished, implementing), nouns (requirements, grades, test scores, lottery system), adjective (based), and adverbs (primarily, instead)].
- Schools that "repeatedly distort or misrepresent verifiable historical facts" or "omit relevant and important context" or "advertise or promote ideologies or socio-political causes or organizations" could face a loss of state funding, state accreditation or tax-exempt status.** [Article 2, lexical density = 67.57%, the lexical items in this sentence were nouns (schools, facts, context, ideologies, causes,

- organizations, loss, state funding, state accreditation, status), adverb (repeatedly), verbs (distort, misrepresent, omit, advertise, promote, face), and adjectives (verifiable, historical, relevant, important, sociopolitical, tax-exempt)].
5. **Critics say this language is so broad as to effectively outlaw any discussion of L.G.B.T. people in elementary school classrooms, or at the very least, strongly discourage teachers from raising those issues, regardless of context.** [Article 2, lexical density = 65.79%, lexical items used in this sentence were nouns (critics, language, discussion, L.G.B.T. people, school, classrooms, at least, teachers, issue, context), verbs (say, outlaw, discourage, raising), adverbs (so, effectively, very, strongly), and adjectives (broad, elementary)].
 6. **The most disturbing efforts to monitor schools and teachers for wrong-think involve actual surveillance.** [Article 2, lexical density = 71.43%, lexical items used in this sentence were adverb (the most), adjectives (disturbing, actual), nouns (efforts, school, teachers, wrong-think, surveillance), and verbs (monitor, involve)].
 7. **Founded in 1628, Collegiate has a long list of prominent graduates that includes one of America's founding fathers, New York City's first governor, John F. Kennedy Jr. And the actor David Duchovny.** [Article 3, lexical density = 71.88%, the lexical items used in this sentence were verbs (founded, has, includes), nouns (Collegiate, list, graduates, one, America, fathers, New York City, governor, John F. Kennedy Jr., actor, David Duchovny), and adjectives (long, prominent, founding, first)].
 8. **Most graduates go on to top-tier colleges.** [Article 3, lexical density = 71.43%, the lexical items used in this sentence were adjectives (most, top-tier), nouns (graduates, colleges), and verb (go)].
 9. **The responses ranged from positive (“iconic,” “history,” “ever-present”) to negative (“racism,” “antisemitism,” “embarrassing”).** [Article 3, lexical density = 76.92%, the lexical items used in this sentence were nouns (responses, history, racism, antisemitism), verb (ranged), and adjectives (positive, iconic, ever-present, negative, embarrassing)].
 10. **It's that New York's public schools have catastrophically failed so many students through lowered expectations, diminished curriculums and limited opportunities for accelerated learning despite the highest per pupil spending of the country's largest school systems.** [Article 4, lexical density = 68.57%, the lexical items used in this sentence were nouns (New York, schools, students, expectations, curriculums, opportunities, learning, pupil, country, school systems), adjectives (public, many, diminished, limited, accelerated, the highest, largest), adverbs (catastrophically, so,), and verbs (failed, lowered, spending)].
 11. **At Brooklyn Tech, Powell describes a “river of teenagers” who are “Bengali and Tibetan, Egyptian and Chinese, Sinhalese and Russian, Dominican and Puerto Rican, West Indian and African American.”** [Article 4, lexical density = 65.52%, the lexical items used in this sentence were nouns (Brooklyn Tech, Powell, river, teenagers), verb (describes), and adjectives (Bengali, Tibetan, Egyptian, Chinese, Sinhalese, Russian, Dominican, Puerto Rican, West Indian, African American)].
 12. **In 1981, Powell reports, nearly two-thirds of Brooklyn Tech's students were Black or Latino.** [Article 4, lexical density = 64.29%, the lexical items used in this sentence were nouns (Powell, two-thirds, Brooklyn Tech, students), verb (reports), adverb (nearly), and adjectives (Black, Latino)].
 13. **Senator Rand Paul, who has opposed vaccine mandates, called Omicron “nature's vaccine.”** [Article 5, lexical

density = 83.33%, the lexical items used in this sentence were nouns (Senator Rand Paul, vaccine mandates, Omicron, nature's vaccine), and verb (opposed)].

14. Also, kids with chickenpox require miserable prolonged isolation at home. [Article 5, lexical density = 80%, the lexical items used in this sentence were adverb (also), nouns (kids, chickenpox, isolation, home), verb (require), and adjectives (miserable, prolonged)].

15. States and school districts that have announced plans for school vaccination requirements already face backlash. [Article 5, lexical density = 73.33%, the lexical items used in this sentence were nouns (states, school districts, plans, school vaccination requirements, backlash), verbs (announced, face), and adverb (already)].

difficult to read, and best understood by college graduates. Meanwhile, according to Flesch-Kincaid grade level, the grade level of these five articles was around 12 to 15, indicating that they were on par with texts readable by at least college students. More information were provided in **Table 4**.

The average sentence length in the five news text samples was about 22 to 30 words. If we calculated them, we would find the average sentence length of these five articles was 25.64 words. This corresponded to a good average of sentence length, which must be around 20-25 words. Furthermore, the news articles from The New York Times which were taken as the samples used more monosyllabic words, the percentage of monosyllabic words used in these five articles was about 59% to 63%, and the average was 61.8%. But, they used fewer polysyllabic words, particularly trisyllabic and longer words. Overall, the shorter the sentence length and shorter words used in the articles reduced the readability. Then, the lower the score of readability made the higher grade level, which meant the text would be more difficult to read.

Table 5 provided a summary of the findings from the analysis of news articles 1, 2, 3, 4, and 5. The conclusion as follow:

1. Each news article had a different number of lexical density percentage. According to Ure's (1985) lexical density theory, a text is classified as written language if its lexical density exceeds 40%. All of the five news articles had a lexical density percentage that exceeded 40%, they were even more than 54%, according to the analysis, indicating that the five news articles were all appropriate for written texts. The analysis of the five examined news articles revealed that the higher lexical density percentage a text had, the more information the text could convey.

Table 4. Readability Analysis

	Ar 1	Ar 2	Ar 3	Ar 4	Ar 5
Average sentence length	22.94	30.56	23	27.52	24.2
Median sentence length	24	31.5	23	27	24
Standard deviation of sentence length	9.8	13.96	11.67	13.68	11.55
Average number of syllables per word	2	2	2	2	2
Percentage of single syllables in text	59%	61%	63%	63%	63%
Percentage of double syllables in text	23%	21%	20%	21%	22%
Percentage of 3+syllables in text	19%	18%	17%	16%	15%
Flesch Reading Ease Score	38.1	37.3	46	40.9	46.8
Flesch-Kincaid Grade Level	13.7	14.9	12.46	14.5	12.8
Readability	DtR	DtR	DtR	DtR	DtR
Grade Level	Col	Col	Col	Col	Col

DtR : Difficult to read
Col : College

Readability Analysis

Based on the results, the reading level of the five articles was around 37 to 46, according to the Flesch Reading Ease Score, if the text score was around 30 to 50, then the text was

Table 5. Lexical density and readability analysis of the news article samples

No	Title	LD (%)	FrS	FKG
1.	San Francisco Ousts School Board Members	58.41	38.1	13.7
2.	You Just Can't Tell the Truth about America Anymore	56.97	37.3	14.9
3.	The Battle over Race, Tradition, and an Elite Private School's Mascot	57.39	46	12.46
4.	The Lessons of Brooklyn Tech	54.72	40.9	14.5
5.	Polio, Chickenpox, Measles ... Covid-19? The Case for Vaccine Mandates in Schools.	55.02	46.8	12.8

LD : Lexical Density

FrS : Flesch Reading Ease Score

FKG : Flesch-Kincaid Grade Level

- Every text had a different amount of Flesch reading ease score. In response to Flesch's (1948) claim that the lower the score is, the more difficult the text is to read. If the score is lower than 50, the text is difficult to read. The analysis revealed that all five news articles had low Flesch reading ease score, they were even lower than 47, which was caused by the sentence length and number of syllables in the texts. The analysis demonstrated that the longer the sentences in the texts were, the more complicated the text was to read.
- The Flesch-Kincaid Grade Level varied among the texts. Flesch-Kincaid Grade Level corresponds to the educational grade level in the US. It displays the level of education of readers who can comprehend a text. In response to Flesch's (1948) assertion that the difficulty of reading a text based on the readers' educational level or age increases with score. A score of 12.5 to 15 indicates that the text is challenging to read, so the readers need to be at least in the collegiate level or aged 18 years old to comprehend the text they are reading. The analysis showed that all five news articles had a score of 12.5 and higher. Based on

the result of analysis, the five news articles could only be read by collegiate students and readers aged 18 years old or more. According to the analysis, the level of education was correlated to the readers' reading comprehension.

It had been discovered that the news articles in The New York Times were mostly difficult to understand for readers under 18 years old. The articles were best understood by the college students or graduates, because they had high percentage of lexical density (which was around 54% to 58%), low score of readability from Flesch Reading Ease Score (which was around 37 to 46), and the grade level was for collegiate level (which was around 12.5 to 15). The lexical density of the five articles was much higher than the lowest score for ordinary written materials and scholarly writings. The high percentage of nouns in the news articles from The New York Times may be due to the unique terms or people's names in the articles that made them have high lexical density.

In terms of text readability, the readability average of the five news articles according to the Flesch Reading Ease was 41.82, which meant the texts were difficult to read. This was collegiate level, which was around 30 to 50. The grade level average of the five news articles according to Flesch-Kincaid Grade Level was 13.67, which was equal to texts readable by college students or graduates. The low readability index of the articles from The New York Times could be linked to the high percentage of monosyllabic words at the expense of polysyllabic words, as well as the high number of words per sentence (with the average of 25.64 words). The amount of words each sentence was fairly high, and it was near to the maximum in published writings.

The results here indicated that qualitative analysis of written texts, in this case news articles, utilizing online software appeared to be a worthwhile effort. The websites were very helpful for the researchers in examining the data and filling in the tables. This

technology might tell the difference between written texts generated by professionals or not. They also would help the readers to find the texts they could understand the best. Using such tools to evaluate the written texts might be advantageous for the readers and made sure if The New York Times could be used as one of credible and reliable sources or online newspapers. More research, however, was needed to determine how reliable the online text analyzer would be in assessing the written texts.

Conclusion

As a result of these findings, it can be concluded that The New York Times can be used as one of the reliable and credible sources for getting information because the five news article samples have high lexical density percentage. It is based on the Ure's and Halliday's lexical density theories; the written text must have at least 40% of lexical density percentage. The article samples have a percentage of lexical density of 58.41% for article 1, 56.97% for article 2, 57.39% for article 3, 54.72% for article 4, and 55.02% for article 5. Meanwhile, the readability of the five news item samples is evaluated by using Flesch Reading Ease Score and Flesch-Kincaid Grade Level. The material is difficult to read if the Flesch reading ease score is lower than 50. The news article samples also have varying Flesch-Kincaid Grade Levels. The Flesch-Kincaid Grade Level reveals the degree of education of readers who are able to understand a text. The score increases as on the reader's age or educational background. The reading ease score and grade level in the article samples are 38.1 and 13.7 for article 1, 37.3 and 14.9 for article 2, 46 and 12.46 for article 3, 40.9 and 14.5 for article 4, and 46.8 and 12.8 for article 5. It can be inferred from these results that majority of the news articles in The New York Times are difficult to read for readers under the age of 18. The articles have a low reading ease score (between 37 and 46) and a grade level appropriate for collegiate level, making

them best comprehended by college students (which is around 12.5 to 15).

References

- Alvermann, D. E., dan Hague, S. A. (1989). Comprehension of counterintuitive science text: Effects of prior knowledge and text structure. *Journal of Educational Research*, 82(4), 197–202.
- Amer, M. A. B., & Baarah, H. A. (2021). Lexical density and readability of secondary stage English textbook in Jordan. *International Journal for Management and Modern Education*, 2(2), 11-20. <https://doi.org/10.33258/birle.v4i3.2438>
- Anderson, M., & Anderson, K. (1997). *Text types in English*. MacMillan Education.
- Biagi, S. (2012). *Media impact: An introduction to mass media* (10th ed.). Wadsworth; Cengage Learning.
- Boykoff, M. T., & Yulsman, T. (2013). Political economy, media, and climate change: Sinews of modern life. *Wiley Interdisciplinary Reviews: Climate Change*, 4(5), 359-371. <https://doi.org/10.1002/wcc.233>
- Britannica, T. Editors of Encyclopaedia (2021, November 5). *The New York Times*. *Encyclopaedia Britannica*. <https://www.britannica.com/topic/The-New-York-Times>
- Din, M., & Ghani, M. (2020). Corpus-based study of identifying verb patterns used in Pakistani newspaper headlines. *Theory and Practice in Language Studies*, 10(2), 149-156. <https://dx.doi.org/10.17507/tpls.1002.02>
- Djuharie, O. S. (2008). *Genre text*. Yrama Widya.
- Dubay, H. W. (2004). *The principle of readability*. Costa Masa.
- Dubay, H. W. (2007). *Smart language: Readers, readability, and the grading of text*. Costa

- Masa.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221-233. <https://doi.org/10.1037/h0057532>
- Gerot, L., & Wignell, P. (1994). *Making sense of functional grammar: An introductory workbook*. Antipodean Educational Enterprises.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. Edward Arnold.
- Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. The Falmer Press.
- Hanafiah, R., & Yusuf, M. (2016). Lexical density and grammatical intricacy in linguistic thesis abstract: A qualitative content analysis. *Proceedings EEIC*, 1(2), 43-46.
- Jaung, W., & Carrasco, L. R. (2022). A big-data analysis of human-nature relations in newspaper coverage. *Geoforum*, 128(4), 11-20. <https://doi.org/10.1016/j.geoforum.2021.11.017>
- 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.
- Madinah, N. (2018). *Lexical metaphor in headlines of The Jakarta Post*. [Bachelor's thesis, Universitas Muhammadiyah Sumatera Utara]. UMSU Respository. <http://repository.umsu.ac.id/handle/123456789/3856>
- Qub'a, A. A., Guba, M. N. A., & Hasan, G. A. (2022). Text readability and lexical density of written text among EFL learners. *Journal of Language and Linguistic Studies*, 18(1), 959-969. <https://doi.org/10.52462/jlls.231>
- Richard, J. C. (1992). *Language teaching and applied linguistics*. British Library.
- 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: Jurnal Pendidikan*, 14(1), 741-752. <https://doi.org/10.35445/alishlah.v14i1.1157>
- Rudell, M. R. (2001). *Teaching content reading & writing*. John Wiley & Sons, Inc.
- Sari, D. (2016). Measuring quality of reading materials in English textbooks: The use of lexical density method in assessing complexity of reading materials of Indonesia's curriculum-13 (k-13) English textbook. *Journal of Applied Linguistics and Literature*, 1(2), 30-39.
- Soyibo, K. (1996). A comparison of communication strategies among three Caribbean high-school biology textbooks. *Journal of Biology Education*, 30(3), 190-194.
- Ure, J. (1971). Lexical density and register differentiation. In G. Perren & J. L. M. Trim (Eds.), *Applications of Linguistics*, London: Cambridge University Press, 443-452.
- Wan, S. (2020). The New York Times and China Daily Online: Linguistic complexity in news headlines. *European Journal of Literature, Language, and Linguistics Studies*, 4(3), 43-75. <http://dx.doi.org/10.46827/ejll.v4i3.217>
- Woods, B., Moscardo, G., & Greenwood, T. (1998). A critical review of readability and comprehensibility test. *The Journal of Tourism Studies*, 9(2), 49-61.
- Yang, S., Zhang, W., & Yuan, Z. (2021). Media reports of the COVID-19 pandemic: A computational text analysis of English reports in China, The UK, and The US. *Scientific Research Publishing*, 9(1), 47-62. <https://doi.org/10.4236/ajc.2021.92004>