

Lexical density in English newspapers – a cross-analysis of *the New York Times* and *Arab News*

Saudi Journal of
Language Studies

65

Hayat Alzahrani
Najran University, Najran, Saudi Arabia and
King Khalid University, Abha, Saudi Arabia

Received 12 December 2024
Revised 15 January 2025
Accepted 20 January 2025

Abstract

Purpose – This study examines the lexical density of articles from *The New York Times (TNYT)* and *Arab News (AN)*, analyzing its impact on the readability and accessibility of content for general audiences, particularly non-native English speakers. It also considers other features affecting readability in these newspapers.

Design/methodology/approach – The chosen method for lexical density is Ure's content-to-total word ratio, while readability is calculated with the help of the Flesch–Kincaid Grade Level. The analysis involves the use of descriptive statistics and Pearson's correlation coefficient to determine the relationship between lexical density and readability. A qualitative content analysis is used for in-depth analysis of contextual factors affecting text comprehensibility. The dataset comprises 100 articles: 50 articles from *TNYT* and 50 articles from *AN*, between 2022 and 2024 and covered topics related to climate change and the environment.

Findings – The findings revealed that *AN* has an average lexical density of 56.54% while *TNYT* has 57.96%. Nevertheless, the above difference is not statistically significant. In addition, lexical density has a positive correlation (0.6332) with Flesch–Kincaid Grade-Level scores. Therefore, this implies that reading comprehension decreases as lexical density increases. Qualitative findings indicate that *AN* emphasizes regional adaptation and increased contextual explanations, which can increase comprehensibility for non-native English speakers.

Research limitations/implications – One of the limitations of this research involves the exploration of two English-language newspapers from Western and Middle-Eastern world regions. This means that the findings cannot accurately reflect the audiences in these two regions. Additionally, this study did not involve the comparison of tolerance thresholds for lexical density among different groups, such as native and non-native English language speakers. In the future, researchers should examine the same topic by considering other Western and Middle-Eastern newspapers that were not covered in this study. In addition, interviews with experts and tests should be administered to compare the impacts of lexical density on comprehensibility between native and non-native English language speakers.

Originality/value – This research contributes to the knowledge of how lexical density affects readability in Western and non-Western English newspapers. It provides insights for media outlets aiming to enhance content accessibility for diverse, global audiences.

Keywords Lexical density, Readability, Comprehensibility, Linguistic complexity, English-language newspapers

Paper type Research paper

English-language newspapers serve many individuals across the world. The media significantly contributes to disseminating information on human relationships, raising awareness of pressing concerns, and influencing consumption behavior changes. The increased digitization of the globe allows individuals to search for information from online media outlets. Newspapers, such as *The New York Times (TNYT)*, published in the United States, and *Arab News (AN)*, representing the Saudi-Arabian population, disseminate information in English. Nevertheless, lexical density, defined as the ratio of content words to the overall words, has emerged as a crucial factor in determining the readability and complexity of a discourse (Bakuuro, 2024;



© Hayat Alzahrani. Published in *Saudi Journal of Language Studies*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Saudi Journal of Language Studies
Vol. 5 No. 2, 2025
pp. 65-75
Emerald Publishing Limited
e-ISSN: 2634-2448
p-ISSN: 2634-243X
DOI 10.1108/SJLS-12-2024-0070

[Khorina and Handani, 2022](#)). The variation in lexical density can create barriers to the overarching aim of media outlets to inform and engage diverse sociodemographic groups. Previous research has indicated that higher lexical density increases cognitive load and reduces reading speed ([Liu and Dou, 2023](#)). This study examines the lexical density of *TNYT* and the *AN* articles published between 2022 and 2024. The focus of this study is significant given the two publications' different cultural and socio-political contexts and the increased adoption of English as a global language. In addition, the comparative approach is relevant since the *AN* serves a predominantly English as a second language (ESL) audience. In addition, research has not explored lexical density, readability, and comprehensibility of newspapers in cross-cultural settings. This study can suggest strategies for enhancing the accessibility of news content to diverse audiences.

Research questions

This study explores three research questions:

- RQ1. How does lexical density differ between articles published between 2022 and 2024 in *TNYT* and the *AN*?
- RQ2. What is the impact of the lexical density of these articles on the comprehensibility and readability by the general audience?
- RQ3. What other factors other than lexical density affect the comprehensibility and readability of the articles in the two newspapers?

Originality and contribution

This research is the first attempt to compare the lexical density and the overall readability of newspapers representing two different cultures, *The New York Times* and *Arab News*. Therefore, it reveals patterns in cross-cultural journalism that can guide organizations in adjusting their writing patterns for increased engagement of audiences in the global segment. Moreover, the results can help news entities in the development of effective strategies to enhance the readability and comprehensibility of articles without sacrificing their information-carrying capacity.

Definition of terms

Lexical density: The proportion of content words relative to the total word count in a text or corpus ([Ure, 1971](#)).

Content words: Nouns, verbs, adjectives, and adverbs that carry meaning ([Ure, 1971](#)).

Readability: The ease with which an individual can read and understand a material ([Kembaren and Aswani, 2022](#)).

Comprehensibility: The degree to which a text can be understood by its intended audience ([Kembaren and Aswani, 2022](#)).

Lexical items: Individual words that have semantic content and contribute to the overall lexical density in a text ([Khorina and Handani, 2022](#)).

Literature review

Lexical density indicates the information-carrying capacity in written and oral discourse. Lexical density is measured as the proportion of content words, which primarily encompass adverbs, verbs, nouns, and adjectives, in the corpus or overall text ([Ure, 1971](#)). In other approaches for computing lexical density, lexemes are expressed as a ratio of a ranking clause (words that convey a complete idea) ([Halliday, 1985](#)). Most studies rely on [Ure's \(1971\)](#)

method in which the functional perspective of language is de-emphasized, and multiple words are not integrated into a single unit (Aziz and Riaz, 2024). Ure's (1971) method is more straightforward and applies primarily to linguistic appraisal.

Lexical density in news media

Several studies have examined lexical density in spoken and print media outlets. Examination of *TNYT*, *Business of Fashion*, and *BBC News* using Halliday's (1985) and Ure's (1971) approaches led to lexical density scores above 50% (Kembaren and Aswani, 2022; Khorina and Handani, 2022; Rahayu and Syaifullah, 2022). Khorina and Handani (2022) used a sample of 10 pieces of written and spoken news from *BBC News*. The findings indicate that written news had slightly higher average lexical density (55.05%) compared to spoken samples (52.15%). Nevertheless, research has indicated that articles in magazines vary significantly depending on the topic and target audience. In their research, Sari and Ekawati (2021) found that the lexical density for articles in *Reader's Digest Magazine* covering various topics was 64.5% for travel, 54.68% for parenting, 58.84% for language, 61.56% for animals, and 61.34% for health. The findings indicate that the majority of the articles could be classified as intricate since their lexical density values were above 60%.

Readability and comprehension

A paucity of research has established a relationship between lexical density, readability, and comprehension in different contexts. Bakuuro (2024) applied both Flesch's (1948) and Gunning's (1952) frameworks to examine readability of high school English texts. The authors inferred a negative correlation between lexical density and readability. Both Ure's (1971) and Halliday's (1985) approaches provided closer values of linguistic complexity. Liu and Dou (2023) indicate that higher ratios of lexemes are associated with a greater cognitive load during the processing of the information. In another study (Rahayu and Syaifullah, 2022), *TNYT* articles had a lexical density of 58.1% congruent with a level 12 of grammatical intricacy, compared to the *Business of Fashion's* lexical index of 54.2% and corresponding grammatical intricacy level of 8. Nevertheless, the above study did not examine the relationship between readability and lexical density. While widely used in the literature, readability formulas have been criticized for focusing on a limited number of textual features to inform reader comprehension (Crossley, 2024). Studies have established additional factors that influence reader comprehension. For example, Ji *et al.* (2023) demonstrated that information distribution over long documents influence comprehension. In addition, Hackemann *et al.* (2022) indicated that technical terms and longer sentences with multiple clauses can substantially reduce the comprehensibility of a corpus.

Cross-cultural considerations

Studies have examined English-language newspapers in ESL countries, revealing important cultural variations (Ghani *et al.*, 2022; Liu and Dou, 2023; Salihoglu and Karatepe, 2023). Using Halliday's (1985) model of systemic functional linguistics, Ghani *et al.* (2022) found that the ratios of lexemes in Pakistani and UK news articles were 49.89 and 53.09%, respectively. The findings suggest that cultural adaptations in non-native English-speaking countries often lead to simplified syntactic structures. Liu and Dou (2023) found that translation processes in cross-cultural contexts impact lexical complexity. They found that direct expression of content in English was linked with more complex vocabulary and content words compared to the interpretation of texts from Chinese or Russian to English.

Research gap and justification of research objectives

Some studies have examined lexical density indices in Western newspapers such as *BBC News* and *TNYT* (Kembaren and Aswani, 2022; Khorina and Handani, 2022; Rahayu and

Syaifullah, 2022; Sari and Ekawati, 2021). However, limited amount of research that examined Western and Middle-Eastern English-language newspapers did not consider the relationship between information density and corresponding readability levels (Ghani *et al.*, 2022; Salihoglu and Karatepe, 2023). Therefore, this study fills this gap by comparing the lexical density and readability of articles from different cultural contexts, Western and Middle-Eastern publications. To address these gaps, this study is structured into three objectives:

- (1) To examine the difference between lexical density of articles published in *TNYT* and *AN*.
- (2) To investigate the impact of lexical density on readability and comprehensibility.
- (3) To identify factors beyond lexical density that affect readability.

The examination of these objectives will address the gap in the role of lexical and structural choices in cross-cultural news. In addition, the objectives provide a framework for understanding how English-language newspapers can adjust and present their messages in a manner that aligns with the diverse international audiences. The objectives support a study approach that combines quantitative and qualitative approaches, which contributes to a better understanding of textual complexity, as well as other cultural-linguistic aspects influencing the reader's comprehension. The findings are relevant to news agencies aiming to disseminate detailed informational content while ensuring that their articles demonstrate reasonable comprehensibility and readability for both the local and the international audiences.

Methodology

Design

This study applied a mixed-method approach to examine the lexical density and readability of *TNYT* and *AN* newspapers. The quantitative method involves analyzing numerical data to elucidate facts about the phenomenon associated with the population. In addition, it reduces bias associated with subjective interpretation of results inherent to qualitative designs. A quantitative approach is relevant given that the current study involved comparing linguistic measures for the two newspapers. A qualitative content analysis allowed for in-depth analysis of contextual factors affecting text comprehensibility.

Data collection

Sampling. The sample in this study involved 100 newspaper articles published between 2022 and 2024 and covered topics related to climate change and environment. Fifty articles were obtained from *TNYT* and the remaining ones from *AN*. A purposive sampling was applied to help the researcher select articles that are relevant to the study objectives. The sampling approach can enhance the internal validity of findings by focusing on participants who truly represent the phenomenon under investigation. The method was useful for the collection of specific and meaningful data related to lexical density in media discourse.

Data sources and selection criteria. Articles were retrieved from the official websites of *TNYT* and *AN*. The selection criteria involved English-language articles and a minimum of 300 words. In addition, articles from wire services were excluded from the study.

Quantitative analysis

Lexical density measurement. Ure's (1971) formula was applied to evaluate the lexical density of words within the selected articles. The content-to-total word ratio indicates the percentage of nouns, verbs, adjectives, and (lexical items) in the text. Prepositions are usually excluded from the computation due to a lack of semantic content. The formula is given below (Khorina and Handani, 2022):

$$\text{Lexical density} = [(\text{Lexical items})/(\text{total number of words})] * 100$$

Comprehensibility and readability analysis. In addition, the Flesch-Kincaid Grade Level, developed by Kincaid *et al.* (1975), was used to calculate the readability aspect. Readability is defined as the proficiency level required for individuals to comprehend the text; therefore, it was synonymous with comprehensibility (Kembaren and Aswani, 2022). Flesch-Kincaid framework accounts for the sentence length and syllable count to establish the grade level required to comprehend each text. As shown in Table 1, the complexity of the text increases as the score moves from a low value to higher Flesch-Kincaid Grade Level scores (Kincaid *et al.*, 1975). The Automatic Readability Checker, a text analysis software, which analyzes discourses based on Flesch-Kincaid Grade Level, was applied for data entry and processing. Manual verification and adjustment were performed. Data were stored in an Excel spreadsheet for further statistical analysis. Flesch-Kincaid formula for appraisal of readability is as follows (Kincaid *et al.*, 1975):

$$0.39 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 11.8 \left(\frac{\text{total syllables}}{\text{total words}} \right) - 15.59$$

Statistical analysis. Descriptive statistics were completed to determine the means and standard deviations for lexical density and Flesch-Kincaid grade-level scores. The impact of lexical density on readability scores was examined through the computation of Pearson's correlation coefficient. In addition, independent *t*-tests were conducted to examine if the mean lexical density of the two newspapers varied significantly.

Qualitative content analysis

The qualitative content analysis involved selecting five articles from each of the two newspapers. A thorough examination of the articles is essential for familiarization and understanding assumptions and patterns in the content (Nicmanis, 2024). Sentences and phrases from the articles were analyzed to determine cultural-linguistic factors that can affect the comprehensibility for general audience. The researcher formed categories and themes that allowed the identification of the similarities and differences between *AN*'s and *TNYT*'s climate change reporting styles.

Validity and reliability

Several measures were taken to ensure that the findings reflected the linguistic characteristics of each newspaper. Reliability focuses on the consistency of measures, while validity indicates the accuracy of measures. Ure's (1971) lexical density formula and the Flesch-Kincaid Grade Level (Kincaid *et al.*, 1975) has been proven reliable in previous studies. In addition, two researchers independently measured the variables for a subset of articles. The results indicated a high level of inter-rater reliability. The utilization of linguistic measures from a respected

Table 1. Interpretation of Flesch-Kincaid grade level

Flesch-Kincaid score	School level	Age range	Readability ease
1–5	1st–5th Grade	11	Very Easy
6	6th Grade	11–12	Easy
7	7th Grade	12–13	Fairly Easy
8–9	8th–9th Grade	13–15	Standard; Plain English
10–11	10th–11 Grade	15–17	Fairly Difficult
12–15	College	17–20	Difficult
16+	College Graduates	20+	Very Difficult

Source(s): Adapted from Kembaren and Aswani (2022) and Kincaid *et al.* (1975)

scholar, Ure (1971), and unambiguous operationalization of readability and lexical density enhanced the content and construct validity of the study. Nevertheless, external validity might be limited to newspapers with audiences and styles similar to those employed in this research.

Results

The sample for this study comprised 100 articles from *TNYT* and *AN*. Table 2 shows the descriptive statistics of the lexical density for the two newspapers. The summary of the individual lexical density elements for each newspaper is shown in Table 3. The average readability values in the form of Flesch-Kincaid Grade Level for *TNYT* and *AN* are depicted in Table 4. In addition, this study involved testing the statistical significance of the difference between *TNYT* and *AN* lexical density, which is shown in Table 5. A correlational analysis of the lexical density and Flesch-Kincaid Grade Level scores of all the articles from two newspapers is shown in Table 6.

Qualitative analysis results

Regional adaptation

Arab News extensively relied on local context examples, especially from the Middle East and North Africa (MENA) region. The publication’s staff referenced detailed country-specific carbon capture and renewable initiatives. Articles focused on Saudi Arabia’s projects, MENA

Table 2. Lexical density results for *TNYT* and *AN* articles

Elements	<i>TNYT</i>	<i>AN</i>
Mean lexical density	57.96%	56.54%
Standard deviation	4.00	3.55
Range	50.12–63.36%	51.26–63.96%
Median	59.08%	56.62%

Source(s): Data based on author’s analysis of *AN* and *TNYT* articles

Table 3. Comparison of average values for various lexical density elements between the two newspapers

Elements	<i>TNYT</i>	<i>AN</i>
Nouns	35.32%	32.43%
Verbs	13.98%	10.74%
Adjectives	7.21%	9.23%
Adverbs	3.45%	4.14%

Source(s): Data based on author’s analysis of *AN* and *TNYT* articles

Table 4. Flesch-Kincaid grade level results for *TNYT* and *AN* articles

Newspaper	<i>TNYT</i>	<i>AN</i>
Mean grade level	13.45	12.90
Standard deviation	1.82	1.32
Range	11.83–16.69	10.53–16.87
Median	12.95	12.75

Source(s): Data based on author’s analysis of *AN* and *TNYT* articles

Table 5. *t*-test results for the difference between *TNYT* and *AN* lexical density

<i>t</i> -test: two-sample assuming unequal variances		
	<i>TNYT</i> lexical density	<i>AN</i> lexical density
Mean	57.9634	56.5356
Variance	16.02312902	12.63464963
Observations	50	50
Hypothesized mean difference	0	
df	97	
<i>t</i> -stat.	1.885954096	
$P(T \leq t)$ one-tail	0.031146508	
<i>t</i> -critical one-tail	1.66071461	
$P(T \leq t)$ two-tail	0.062293016	
<i>t</i> -critical two-tail	1.984723186	

Source(s): Data based on author's analysis of *AN* and *TNYT* articles

Table 6. Correlation analysis between lexical density and readability (Flesch-Kincaid grade level) for both newspapers

	Lexical density	Flesch-Kincaid grade level
Lexical density	1	
Flesch-Kincaid grade level	0.633233065	1

Source(s): Data based on author's analysis of *AN* and *TNYT* articles

as a food deficit economy, and the region's potential water scarcity. The analysis revealed more global references with less contextual details in *TNYT*.

Technical content presentation

Arab News and *TNYT* demonstrated different strategies for presenting technical content regarding climate change issues. *Arab News*, for instance, provided explicit definitions or analogies for terms such as voluntary carbon market and circular carbon initiative. On the other hand, *TNYT* writers mostly assumed that their readers had prior knowledge of environmental discourse and international policy frameworks. *The New York Times* used unexpanded acronyms and did not include explanatory phrasing for scientific terms such as thermal batteries, carbon capture, and clean hydrogen. However, similar to *TNYT*, one of the *AN* articles assumes that the audience has background knowledge about global climate initiatives due to the use of terms such as The Conference of Parties (COP27) in the form of acronyms and without contextual elaboration.

Structural complexity

Structural analysis focused on sentence complexity, paragraph organization, and explanatory transitions. *The New York Times* demonstrated a higher frequency of complex sentences than *AN* (62 vs 54%). On the other hand, *AN* articles had shorter paragraphs (average of 3–4 sentences) than *TNYT* (6–7 sentences). However, both *AN* and *TNYT* showed comparable use of explanatory transitions.

Discussion

RQ1. How does lexical density differ between the articles published between 2022 and 2024 in *TNYT* and the *AN*?

Table 2 shows that *TNYT* has a slightly higher mean lexical density (57.96%) compared to the *AN* sample (56.54%). The above findings align with [Ure's \(1971\)](#) study which established that lexical density for written news exceeded 40%. The results shown in [Table 3](#) indicate that the lexical items that accounted for the greatest proportion of the lexical density in both sets of newspapers were nouns, with *TNYT* having a slightly higher percentage (35.32%) than *AN* (32.43%). Independent *t*-test results ($p > 0.05$) indicate that the difference between lexical density in *TNYT* and *AN* is not statistically significant, as shown in [Table 5](#). Therefore, *TNYT* and *AN* are comparable in terms of information-carrying capacity.

Prior studies have concentrated on lexical density in Western media outlets such as *TNYT* and *BBC News* without comparing the values with Middle-Eastern media ([Khorina and Handani, 2022](#); [Rahayu and Syaifullah, 2022](#)). However, consistent research has established that the typical lexical density for news articles and other publications that use formal writing is 50–60% ([Maufiroh et al., 2024](#); [Ramadhani et al., 2023](#)). However, some *AN* and *TNYT* articles had lexical density scores that exceeded 60%, as depicted in [Table 2](#). According to [Maufiroh et al. \(2024\)](#), values ranging between 61 and 70% are interpreted as high and indicate a text that is difficult to read. Additionally, these above-average lexical density patterns are common in specialized documents assuming that the audience has an in-depth understanding of the subject matter.

RQ2. What is the impact of the lexical density of these articles on the comprehensibility and readability by the general audience?

A correlational analysis was performed to determine the impact of lexical density on readability. As shown in [Table 6](#), lexical density has a positive correlation (0.6332) with Flesch-Kincaid Grade-Level scores (used to measure the readability and comprehensibility of the text in this research). Higher Flesch-Kincaid Grade-Level scores usually indicate that the text is more difficult to read ([Kembaren and Aswani, 2022](#)). Therefore, the correlation analysis indicates that reading comprehension for *TNYT* and *AN* articles decreases as lexical density increases. Information-dense texts require a higher cognitive load for readers ([Liu and Dou, 2023](#)). Akin to the findings in this study, [Bakuuro \(2024\)](#) found that lexical density had a negative relationship with readability. Nevertheless, the researcher focused on academic context rather than the media. [Rahayu and Syaifullah \(2022\)](#) found that *The New York Times* articles had higher grammatical intricacy than *Business of Fashion* due to higher lexical density.

[Table 4](#) shows that the average readability levels of the two media are *TNYT* (13.45 Grade Level) and *AN* (12.90 Grade Level). The scores indicate that the articles can be read and understood by individuals with at least college education. In addition, *AN* and *TNYT* articles can be categorized as difficult to read (see [Table 1](#)). Given the moderate to strong positive correlation established in [Table 6](#), newspapers should consider lowering lexical density as one of the measures to increase readability and make their content suitable for individuals with an average education level.

RQ3. What other factors other than lexical density affect the comprehensibility and readability of the articles in the two newspapers?

The qualitative analysis offers insight into additional factors other than lexical density that can affect the comprehensibility of *AN* and *TNYT* articles. Although the two newspapers have similar lexical density, *AN*'s regional adaptation, technical content explanation, and lower sentence complexity compared to *TNYT* can enhance reader comprehension for non-native English speakers. *Arab News* focuses on MENA's environmental concerns and uses locally relevant examples such as Saudi Arabia's renewable energy initiatives to elaborate on impacts and practical implementation strategies. *The New York Times* reports climate change issues from a global perspective and demonstrates higher technical content with minimal contextual explanation, reflecting its assumption of reader familiarity with environmental discourse. Prior research has established that complex sentence structures and technical vocabulary

reduce the comprehensibility of texts (Crossley, 2024; Hackemann *et al.*, 2022). Therefore, it can be inferred that *TNYT*'s linguistic features can hinder reader comprehension more than *AN*.

Practical implications

This study has established that higher lexical density is associated with lower readability and comprehensibility. Therefore, it is prudent for news organizations to include lexical density in their writing policies. Recommended optimal density values that can allow news articles to balance information-carrying capacity and comprehensibility for the general audience should range from 50% to 60%, as discussed by Maufiroh *et al.* (2024). The print media should use a multi-tiered approach in which a significant proportion of articles are written with low lexical density for the general public and additional in-depth articles are disseminated for individuals proficient in climate change and environmental policy. Also, based on the qualitative results, journalists should increasingly use contextual explanations for technical terms and integrate simple sentence structures and regional adaptation into their articles to reach individuals with average education levels.

Limitations and future research

One of the limitations of this research involves the exploration of two English-language newspapers from Western and Middle-Eastern world regions. This means that the findings cannot accurately reflect the audiences in these two regions. Additionally, this study did not involve the comparison of tolerance thresholds for lexical density among different groups, such as native and non-native English language speakers. In the future, researchers should examine the same topic by considering other Western and Middle-Eastern newspapers that were not covered in this study. In addition, interviews with experts and tests should be administered to compare the impacts of lexical density on comprehensibility between native and non-native English language speakers.

Conclusion

This study aimed to compare the lexical density of *TNYT* and *AN* articles published between 2022 and 2024 and their impact on the readability of the text. Ure's (1971) content-to-total word ratio method and the Flesch-Kincaid Grade Level were used to measure lexical density and readability, respectively. An independent *t*-test, Pearson's correlation coefficient, and qualitative content analysis were performed to answer research questions. The findings indicated a moderate positive correlation (0.6332) between lexical density and the Flesch-Kincaid Grade Level. Therefore, this means that the readability of the text reduces as lexical density increases. *Arab News* and *TNYT* have comparable lexical density averages, which are greater than 50%. The findings underscore the need for news organizations to establish optimal density ranges to disseminate informative content and simultaneously enhance reader comprehension. In addition, local context integration, explanatory phrasing of technical terms, and less complex structures should be considered as additional measures for increasing comprehensibility for audiences. Future research should examine lexical density and comprehensibility of newspapers from multiple cultural contexts.

References

- Aziz, S. and Riaz, F. (2024), "Comparing the two lexical density measures: the case of ten highest impact factor journals in applied linguistics", *Pakistan Journal of Law, Analysis and Wisdom*, Vol. 3 No. 4, pp. 220-229, doi: [10.24071/llt.v26i1.4971](https://doi.org/10.24071/llt.v26i1.4971).

- Bakuuro, J. (2024), "In the belly of text complexity: unravelling the nexus between lexical density and readability", *Athens Journal of Philology*, Vol. 11 No. 3, pp. 1-20, doi: [10.30958/ajp.X-Y-Z](https://doi.org/10.30958/ajp.X-Y-Z).
- Crossley, S.A. (2024), "Developing linguistic constructs of text readability using natural language processing", *Scientific Studies of Reading*, Vol. 7, pp. 1-23, doi: [10.1080/10888438.2024.2422365](https://doi.org/10.1080/10888438.2024.2422365).
- Flesch, R. (1948), "A new readability yardstick", *Journal of Applied Psychology*, Vol. 32 No. 3, pp. 221-233, doi: [10.1037/h0057532](https://doi.org/10.1037/h0057532).
- Ghani, F., Saleem, T., Majeed, S., Batool, R. and Aslam, M. (2022), "A corpus-based comparative ideational meta-functional analysis of Pakistani English and UK English newspaper editorials on COVID-19", *Cogent Arts and Humanities*, Vol. 9 No. 1, 2114619, doi: [10.3389/fpsyg.2022.800983](https://doi.org/10.3389/fpsyg.2022.800983).
- Gunning, R. (1952), *The Technique of Clear Writing*, McGraw-Hill, New York.
- Hackemann, T., Heine, L. and Höttecke, D. (2022), "Challenging to read, easy to comprehend? Effects of linguistic demands on secondary students' text comprehension in physics", *International Journal of Science and Mathematics Education*, Vol. 20 No. 1, pp. 43-68, doi: [10.1007/s10763-022-10306-1](https://doi.org/10.1007/s10763-022-10306-1).
- Halliday, M.A.K. (1985), *An Introduction to Functional Grammar*, Edward Arnold, London.
- Ji, S., Sun, W. and Marttinen, P. (2023), "Content reduction, surprisal, and information density estimation for long documents", *Frontiers in Psychology*, Vol. 13, 800983, available at: <https://arxiv.org/pdf/2309.06009>
- Kembaren, F.R. and Aswani, A.N. (2022), "Exploring lexical density in the New York Times", *Journal of English Literature, Language and Teaching*, Vol. 7 No. 2, pp. 109-119, doi: [10.32528/ellite.v7i2.8795](https://doi.org/10.32528/ellite.v7i2.8795).
- Khorina, M. and Handani, W.Z. (2022), "Comparing the lexical density of spoken and written news", *Linguistics and ELT Journal*, Vol. 10 No. 2, pp. 18-24, doi: [10.31764/leltj.v10i2.10226](https://doi.org/10.31764/leltj.v10i2.10226).
- Kincaid, J.P., Fishburne, R.P., Rogers, R.L. and Chissom, B.S. (1975), "Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel", Research Branch Report 8-75, Naval Air Station Memphis, TN.
- Liu, Z. and Dou, J. (2023), "Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: a cognitive perspective", *Frontiers in Psychology*, Vol. 14, 1276705, doi: [10.3389/fpsyg.2023.1276705](https://doi.org/10.3389/fpsyg.2023.1276705).
- Maufiroh, Y.L., Hussen, B.T.E. and Femilia, P.S. (2024), "Lexical density of reading materials in an English textbook: a content analysis", *Journal of Language Intelligence and Culture*, Vol. 6 No. 2, pp. 165-174, doi: [10.35719/jlic.v6i2.104](https://doi.org/10.35719/jlic.v6i2.104).
- Nicmanis, M. (2024), "Reflexive Content Analysis: an approach to qualitative data analysis, reduction, and description", *International Journal of Qualitative Methods*, Vol. 23, pp. 66-83, doi: [10.1177/16094069241236603](https://doi.org/10.1177/16094069241236603).
- Rahayu, M.S. and Syaifullah, A.R. (2022), "Comparison of lexical density and grammatical intricacy in news text of readability", *International Journal of Science and Applied Science*, Vol. 6 No. 1, pp. 30-41, doi: [10.20961/ijscs.v6i1.69933](https://doi.org/10.20961/ijscs.v6i1.69933).
- Ramadhani, R., Aulawi, H. and Ulfa, R.L. (2023), "Readability of reading texts as authentic materials: a systemic functional perspective", *Journal of English Language Teaching and Applied Linguistics*, Vol. 8 No. 2, pp. 149-168, doi: [10.30595/lks.v15i2.11102](https://doi.org/10.30595/lks.v15i2.11102).
- Salihoglu, S. and Karatepe, C. (2023), "Investigation of online news about Ukraine-Russia war: a corpus-based critical discourse analysis", *Journal of Language and Literature Research*, Vol. 3 No. 33, pp. 1171-1186, doi: [10.29000/rumelide.1279172](https://doi.org/10.29000/rumelide.1279172).
- Sari, A.N. and Ekawati, R. (2021), "Lexical density in Reader's Digest magazine", *Prosodi*, Vol. 15 No. 1, pp. 17-22, doi: [10.21107/prosodi.v15i1.10483](https://doi.org/10.21107/prosodi.v15i1.10483).

Ure, J. (1971), "Lexical density and register differentiation", in Perren, G.E. and Trim, J.L.M. (Eds), *Applications of Linguistics*, Cambridge University Press, pp. 443-452.

Further reading

Ratajczak, M., Indrarathne, B. and Kormos, J. (2021), "The effects of read-aloud assistance, vocabulary and background knowledge on comprehension of health-related texts of Sri-Lankan English as second language speakers", *Journal of the European Second Language Association*, Vol. 5 No. 1, pp. 133-147, doi: [10.22599/jesla.78](https://doi.org/10.22599/jesla.78).

Corresponding author

Hayat Alzahrani can be contacted at: haalzahrani@nu.edu.sa