

Annotated Bibliography

"Lexical Density in English Newspapers – A Cross Analysis of The New York Times and Arab
-Case Study"

Lexical density has emerged as one of the factors that impact text complexity and the ability of certain audiences to understand the information. The main research question addressed in this annotated bibliography centers on the variation in lexical density in English-language newspapers and its impact on the comprehensibility and readability of these articles by the general audience. In addition, the research question seeks to establish additional factors that influence text complexity in the newspapers. This research is significant because English is being used more globally, and the media needs to write in a way that is both professional and easy to understand for readers from different backgrounds. The bibliography includes studies published between 2020 and 2024, with the exception of seminal studies. The annotation will focus on the key findings, methodology, and the relevance of selected studies to the research objectives.

Bakuuro, J. (2024). In the belly of text complexity: Unravelling the nexus between lexical density and readability. *Athens Journal of Philology*, *11*, 1-20.

<https://doi.org/10.30958/ajp.X-Y-Z>

This study explored how lexical density is associated with complexity and general text readability. The authors applied Gunning's and Flesch's framework to analyze the readability indices of the text. The results indicate that the lexical density values of the text exceeded the recommended minimum of 40 percent. In addition, high lexical density was associated with low readability and vice versa. The study asserts that the reading process forms the basis for receiving information and knowledge from texts. Bakuuro's (2024) work is instrumental to the

current topic since it examines how lexical density impacts audience comprehension. On average, Halliday's and Ure's values were close among the sample texts investigated. This consistency in the approaches used to measure the variables fortifies the reliability of the study. In addition, a quantitative methodology reduces bias associated with the researcher's subjective interpretation of results. The current study concentrated on high school students rather than the general public. However, it complements the information provided by other studies by directly addressing the research question.

Ghani, F., Saleem, T., Majeed, S., Batool, R., & Aslam, M. (2022). A corpus-based comparative ideational meta-functional analysis of Pakistani English and UK English newspaper editorials on COVID-19. *Cogent Arts & Humanities*, 9(1), 2114619.

<https://doi.org/10.3389/fpsyg.2022.800983>

This study employed Halliday's systemic functional linguistic theory to examine the transitivity process and lexical features in the UK and Pakistani newspapers on the coronavirus pandemic published in English. In a sample of 5 UK editorials containing 246 sentences and 5648 words, the lexical density was 53.09 percent, compared to 49.89 percent in Pakistani newspapers. Nevertheless, the material process was the most used transitivity system category followed by the relational process for both editorials. This implies that the texts were objective and unambiguous. In terms of linguistic expression, UK samples had a higher proportion of "circumstance" and "process" components, but a lower percentage of "participants" components than Pakistani editorials. Therefore, UK writers expressed a significant number of lexemes in the form of prepositional phrases, adverbial groups, and verbs compared to Pakistani editors. This study offers adequate evidence since it explored multiple newspapers written in English. In addition, the findings are relevant to the current topic since it examines how editors apply the

transitivity process to develop reasonable meanings and present their views. In addition, the findings will be used to compare the lexical choices in the UK context and Pakistani media discourse. The research extends the topic under interest by providing information that can help individuals make inferences about the expectations of the audiences of the nations.

Ji, S., Sun, W., & Marttinen, P. (2023). Content reduction, surprisal, and information density estimation for long documents. *Frontiers in Psychology, 13*, 800983.

<https://arxiv.org/pdf/2309.06009>

This study presents strategies for enhancing information density in long documents in diverse disciplines such as media (BBC News), healthcare, and movie reviews. The findings indicate that applying attention-based selection and summarization can improve lexical density. Unlike other sources, researchers emphasize that redundant information can increase the reader's cognitive load, lower text processing, and potentially impact classification performance. Information-dense phrases are preferable for rational language users. This study can be applied to the proposed research topic to demonstrate that content reduction strategies can maintain a particular level of lexical density that enhances readability. Nevertheless, these strategies are applicable to clinical contexts, which involve less standard text. News media apply formal language that is more readable than user-generated movie reviews or clinical notes. However, this source is relevant since some machine models improved the lexical richness (density and diversity) of news. Furthermore, the information can be applied to show how media entities can target a particular audience.

Kembaren, F. R., & Aswani, A. N. (2022). Exploring lexical density in the New York Times.

Journal of English Literature, Language and Teaching, 7(2), 109-119.

<http://dx.doi.org/10.32528/ellite.v7i2.8795>

The purpose of this study was to analyze the lexical density of articles published in The New York Times using Halliday and Ure's approaches. In addition, the authors evaluated the readability of respective articles using Flesch's Reading Ease Score. The methodology of the study was primarily qualitative content analysis. The findings suggest that lexical densities were above 50 percent and readability scores were less than 50. This implies that the information could be understood by individuals in grade level 12 and above. A lexical density score of more than 40 percent supports the credibility and reliability of a media outlet in disseminating informative content. A significant strength of this study involves the use of content analysis, which is relevant for analyzing written texts. In addition, the authors used many news articles, although they only concentrated on one media outlet, The New York Times. Nevertheless, the methodology can raise issues about the generalizability of results. This study extends the current topic regarding the impact of lexical density on the audience's understanding of the text.

Khorina, M., & Handani, W. Z. (2022). Comparing the lexical density of spoken and written

news. *Linguistics and ELT Journal*, 10(2), 18-24.

<https://doi.org/10.31764/leltj.v10i2.10226>

This study examined the lexical density of written and spoken news on BBC News. The sample comprised 10 pieces, with 5 written texts being obtained from the BBC's website and the remaining ones from the media outlet's YouTube channel. The average lexical density values for the written and spoken samples were 55.05% and 52.15%, respectively. In addition, the study

established that items that enhanced the density of texts included adverbs, adjectives, nouns, and verbs, although nouns were predominant. The findings suggest that BBC News articles have a greater information-carrying capacity, which surpasses the 40 percent threshold. In addition, this information may indicate an increased cognitive load during the processing of the information. Given that prior sources examined the association between readability and lexical density, this source can be applied to make inferences on the ability of the audience to understand BBC newspapers. Furthermore, the strength of this research involves the use of evidence to support the argument raised. This study is instrumental to the topic under investigation since it assesses the lexical density of the newspapers from UK media.

Liu, Z., & Dou, J. (2023). Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: A cognitive perspective. *Frontiers in Psychology, 14*, 1276705. <https://doi.org/10.3389/fpsyg.2023.1276705>

This study examined the cognitive load associated with lexical features of simultaneously interpreted texts. The researchers considered texts translated from Russian to English (B-to-A), Chinese to English (A-to-B), and the non-interpreted ones (expressed in U.S. English). In this context, A denoted primary language while B referred to secondary language. The findings demonstrated that non-interpreted texts had more complex vocabulary than the B-to-A and A-to-B contexts. Furthermore, B-to-A texts were associated with more simplified lexical features than A-to-B texts. The researchers indicate that simpler lexical characteristics in interpreted articles are linked to interpreters' tendency to minimize their cognitive load. Temporal factors can constrain the working memory and impact the quality of simultaneous interpretation. The researchers did not demonstrate bias in their interpretation of the findings. In addition, they have embraced a balanced approach to support their arguments. This study builds on the current topic

to show that translation affects the lexical density of news articles. In addition, directionality (translation from native to non-native (A-to-B) or secondary to first language (B-to-A) also plays a crucial role. A greater cognitive load is required for A-to-B, thereby resulting in simplified lexical features.

Rahayu, M. S., & Syaifullah, A. R. (2022). Comparison of lexical density and grammatical intricacy in news text of readability. *International Journal of Science and Applied Science*, 6(1), 30-41. <https://dx.doi.org/10.20961/ijscs.v6i1.69933>

This study explored the readability of different newspaper articles in relation to grammatical intricacy (GI) and lexical density (LD). GI was conceptualized as the ratio of clauses to sentences in a text. The New York Times had a GI and LD of level 12 and 58.13% compared to level 8 and 54.18% for Business of Fashion. The authors concluded that the Business of Fashion was easier to comprehend due to lower LD and GI values. Given that the LD average LD value is above 50 percent, the Business of Fashion can provide a high volume of information while catering to individuals with low English proficiency. This source is relevant since it examines newspapers that have not been addressed in other research articles in this annotated bibliography. Additionally, this study used credible information from prior research to back up the inferences that GI can mediate the relationship between LD and comprehension of a text. The findings in this study are instrumental in the proposed topic since they examine lexical profiles in the U.S media. The source can be used for comparative purposes to determine how media outlets balance simplicity and information delivery to suit their audiences.

Salihoglu, S., & Karatepe, C. (2023). Investigation of online news about Ukraine-Russia war: A corpus-based critical discourse analysis. *Journal of Language and Literature Research*, 3(33), 1171-1186. <http://dx.doi.org/10.29000/rumelide.1279172>

This study involved a critical discourse analysis (CDA) of Al Jazeera and BBC news articles about the current Russia-Ukraine war. A corpus-based analysis was employed to determine shared lexemes using frequency tables and concordances. The study found no significant differences between BBC news and Al Jazeera's information density and contextual factors as both entities portrayed Russia as the invader. The articles blamed Vladimir Putin for violating international laws and labeled him as a war criminal. The authors related the BBC News' unfavorable portrayal of Russia to the ideological conflicts between the Western nations and Russia. The strengths of this study center around the examination of a British media outlet with substantial viewership. In addition, the research adds new information on how CDA involving language structures and contextual awareness allows readers to gain an in-depth understanding of the reality of events in the discourse. This study is instrumental to the proposed research topic and complements other sources by examining lexical density using corpus data on an event that has detrimental impacts on human lives.

Sari, A. N., & Ekawati, R. (2021). Lexical density in Reader's Digest Magazine. *Prosodi*, 15(1), 17-22. <http://dx.doi.org/10.21107/prosodi.v15i1.10483>

This study explored the lexical density of multiple articles in Reader's Digest Magazine (RDM). In addition, the computation was based on Ure's model, and Free Claws Web Tagger was used to underline grammatical and lexical items. The results for health, animal kingdom, language, parenting, and travel articles were 61.34%, 61.56%, 58.83%, 54.68%, and 64.05%,

respectively. Subsequently, lexical density scores were used to categorize articles as easy (40%-50%), standard (50%-60%), or hard (60%-70%). The findings suggest that most of the RMD's articles are intricate, requiring individuals with high education levels. The strength of this study involves the use of multiple article types from the RDM (travel, parenting, languages, health, and animals). In addition, the findings are based on the data analysis and can add to the proposed research topic to show lexical density variation across genres. Moreover, the fluctuation can influence how media discourse is structured and understood by audiences. This study extends the proposed topic since it can be utilized to compare lexical density measures with other scholarly works that employed a similar methodology for different news outlets.

Ure, J. (1971). Lexical Density and Register Differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443-452). Cambridge University Press.

This seminal work provides foundational information on lexical density. Ure (1917) defined lexical density as the proportion of content words relative to total word count in a text or corpus. The author emphasized that written language forms tend to have lexical density above 40 percent, and their values are higher than spoken forms. Writers are oriented to use information that is active rather than reflective in nature, compared to the spoken counterpart. In Ure's model, multiple words are not calculated as a single unit. The lexical label content word aligns with the purpose of the proposed study. In addition, the proposed topic will use this seminal work to develop the formula for computing the lexical density of articles from diverse media outlets. This study is relevant to the current research topic since it offers a technical and theoretical framework for analyzing lexical density. Nevertheless, the study does not address how text complexity can impact audience understanding.

Conclusion

The findings consistently indicate that high lexical density was associated with low readability and vice versa. A lower proportion of lexical units in a text reduces the cognitive load for information processing. Additionally, the reviewed works underline the significance of cultural factors, language translation, and sentence length in shaping audience comprehension of text. The interpretation of content from a foreign language to English can result in simpler lexical features and higher readability. However, limited research has explored the relationship between lexical density and reader comprehension of newspapers within cross-cultural settings. Much research has examined newspaper articles published in Western countries, such as BBC News, The New York Times, Business of Fashion, and Reader's Digest Magazine, and their lexical densities are above fifty percent.

References

- Bakuuro, J. (2024). In the belly of text complexity: Unravelling the nexus between lexical density and readability. *Athens Journal of Philology*, *11*, 1-20.
<https://doi.org/10.30958/ajp.X-Y-Z>
- Ghani, F., Saleem, T., Majeed, S., Batool, R., & Aslam, M. (2022). A corpus-based comparative ideational meta-functional analysis of Pakistani English and UK English newspaper editorials on COVID-19. *Cogent Arts & Humanities*, *9*(1), 2114619.
<https://doi.org/10.3389/fpsyg.2022.800983>
- Ji, S., Sun, W., & Marttinen, P. (2023). Content reduction, surprisal, and information density estimation for long documents. *Frontiers in Psychology*, *13*, 800983.
<https://arxiv.org/pdf/2309.06009>
- Kembaren, F. R., & Aswani, A. N. (2022). Exploring lexical density in the New York Times. *Journal of English Literature, Language and Teaching*, *7*(2), 109-119.
<http://dx.doi.org/10.32528/ellite.v7i2.8795>
- Khorina, M., & Handani, W. Z. (2022). Comparing the lexical density of spoken and written news. *Linguistics and ELT Journal*, *10*(2), 18-24.
<https://doi.org/10.31764/leltj.v10i2.10226>
- Liu, Z., & Dou, J. (2023). Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: A cognitive perspective. *Frontiers in Psychology*, *14*, 1276705. <https://doi.org/10.3389/fpsyg.2023.1276705>

Rahayu, M. S., & Syaifullah, A. R. (2022). Comparison of lexical density and grammatical intricacy in news text of readability. *International Journal of Science and Applied Science*, 6(1), 30-41. <https://dx.doi.org/10.20961/ijscs.v6i1.69933>

Salihoglu, S., & Karatepe, C. (2023). Investigation of online news about Ukraine-Russia war: A corpus-based critical discourse analysis. *Journal of Language and Literature Research*, 3(33), 1171-1186. <http://dx.doi.org/10.29000/rumelide.1279172>

Sari, A. N., & Ekawati, R. (2021). Lexical density in Reader's Digest magazine. *Prosodi*, 15(1), 17-22. <http://dx.doi.org/10.21107/prosodi.v15i1.10483>

Ure, J. (1971). Lexical Density and Register Differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443-452). Cambridge University Press.