



RESEARCH ARTICLE

Analysis of lexical richness three book genres based on MSTTR

Atikah Solihah^{1*} & Harini²

¹ Badan Pengembangan dan Pembinaan Bahasa, Jakarta, Indonesia

² Badan Riset dan Inovasi Nasional, Jakarta, Indonesia

*Corresponding Author

atikah.solihah@kemendikdasmen.go.id

ABSTRACT

Lexical richness is one indicator of vocabulary diversity in a text and is important to study in order to understand the characteristics of language across various genres. However, to date, comparisons of lexical richness across books of different genres have been limited. Therefore, this study aims to compare the lexical richness of three Indonesian-language books, namely *Visi Anak Pasar*, *Yang Fana Adalah Waktu*, and *Tata Bahasa Baku Bahasa Indonesia*, using the Mean Segmental Type-Token Ratio (MSTTR) based on lexical richness theory. This study was conducted using a mixed-methods approach, combining quantitative and qualitative methods. The quantitative approach was carried out using an explanatory technique with the MSTTR instrument. Meanwhile, the qualitative approach was carried out using corpus analysis techniques, including frequency analysis using a concordance plot, morphological analysis, and contextual analysis using the Leipzig-Corpora corpus. The results of the study show that the autobiography *Visi Anak Pasar* has the highest average MSTTR value, at 51.5 per cent, followed by the novel *Yang Fana Adalah Waktu* at 45.9 per cent, whilst the book *Tata Bahasa Baku Bahasa Indonesia* has the lowest average MSTTR value, at 39 per cent. These findings suggest that the personal narrative genre exhibits a higher tendency for lexical repetition compared to the other two books. Consequently, MSTTR can be used as a tool for comparing lexical characteristics across genres in a more balanced manner.

Keywords: lexical richness; Mean Segmental Type-Token Ratio (MSTTR); texts genre.

ARTICLE HISTORY

Received : March 15, 2026

Accepted : May 6, 2026

Published : July 24, 2026

LICENSE



Copyright © 2026 The Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.

<https://creativecommons.org/licenses/by/4.0/>

How to Cite This Article

Atikah Solihah & Harini (2026). Analysis of lexical richness three book genres based on MSTTR.

Ghancaran: Jurnal Pendidikan Bahasa dan Sastra Indonesia, 8(1), 303-316.

<https://doi.org/10.19105/ghancaran.v8i1.24236>

INTRODUCTION

The study of written language across various genres is one of the main focuses of modern linguistics, particularly in the fields of corpus linguistics and text analysis. Texts are understood not only as sequences of linguistic forms, but also as a means of representing the author's experiences, ideologies and communicative aims (Hardinanto, 2025). These differences in social function are linked to tendencies in word choice, patterns of repetition and the level of linguistic complexity within each genre (Purba & Sidebang, 2024). Therefore, it is important to analyse the lexical variation in a book in order to understand the linguistic characteristics of a text in a more objective and measurable way (Mastutik & Taufiq, 2021).

Lexical richness refers to the degree of diversity in the vocabulary used in a text (Suhardijanto & Kusumawardani, 2021). Texts with high lexical richness tend to contain a more varied choice of words with a lower level of repetition, whilst texts with low lexical richness exhibit a more limited and repetitive use of vocabulary (Suhendra & Patriani, 2021). Research into lexical richness is important for the development of teaching materials, the assessment of readability, and the mapping of linguistic characteristics across various text genres.

Lexical richness is generally measured using the Type-Token Ratio (TTR), which is the ratio of the number of unique words (types) to the total number of words (tokens) (Suhardijanto & Kusumawardani, 2021). However, TTR has a fundamental weakness in that it is heavily influenced by text length. The longer the text, the more the TTR value tends to decrease, meaning that comparisons between texts become less (Kyle et al., 2021; Savila, 2025). To address these limitations, segment-based measures are used, one of which is the Mean Segmental Type-Token Ratio (MSTTR) (Hardiansyah et al., 2025), namely, calculating the TTR for text segments with the same number of tokens, and then taking the average. Using this procedure, the MSTTR is considered to be more stable and more suitable for comparing the lexical richness of texts of different lengths and genres (McCarthy & Jarvis, 2010).

The MSTTR score will reveal the extent to which a book utilises a varied vocabulary (Muldasari, 2025). Is the vocabulary used in a varied and extensive manner, or is it repetitive and limited? In the context of linguistic research, MSTTR has been applied in various fields, ranging from language acquisition studies, literary stylistics and genre analysis to the assessment of linguistic complexity in textbooks, literary works and academic (Al Pansori et al., 2025; Haryani et al., 2025; Mu'in, 2025; Sapihah et al., 2025).

A number of previous studies support the use of lexical diversity measures in text analysis. MSTTR has been used as a measure of vocabulary variation in narratives by children with special educational needs or autism (Dosi & Anthi, 2025). Covington and McFall (2010) demonstrate that segment-based measures, such as MSTTR, are more reliable than TTR as they are able to reduce text-length bias (Covington & McFall, 2010). In genre studies, Akbarii and Jarvis (2023) demonstrate that lexical diversity can serve as a predictor of genre, as each genre has a distinctive vocabulary profile. Meanwhile, Hanuskova (2025) found a tendency for higher levels of writer proficiency to be associated with greater lexical diversity in second-language learners' texts (Hanušková et al., 2025). These findings indicate that lexical variation is relevant to study from methodological, functional and pedagogical perspectives.

Nevertheless, research on lexical richness in Indonesian-language texts has largely focused on student writing, language acquisition, or the analysis of a specific text type. The methodology also differs, as it utilises TTR. Furthermore, in terms of contribution, much of the research centres on qualitative measurement based on learners' writing. Studies comparing Indonesian-language books across genres using uniform quantitative measurements remain relatively limited. Yet, cross-genre comparisons are crucial for demonstrating how a text's communicative purpose relates to the distribution and variation of vocabulary. It is this gap that forms the basis for the necessity of this study.

The novelty of this research lies in the use of MSTTR to compare three Indonesian-language books from different genres—autobiography, novel and reference book—with data quantities aligned through identical token segmentation. Unlike previous research, which has largely focused on learners' writing or a specific type of text, this study takes cross-genre books as its subject matter, thereby providing a comparative overview of the differences in lexical characteristics across these three genres of Indonesian-language books.

Based on the above discussion, this article aims to compare the lexical richness of three Indonesian-language books representing different genres, namely the autobiography *Visi Anak Pasar* by E. Aminudin Aziz, the novel *Yang Fana Adalah Waktu* by Sapardi Djoko Damono, and the reference book *Tata Bahasa Baku Bahasa Indonesia* by the Language Development and Guidance Agency. These three books each excel within their respective genres. The analysis was conducted using MSTTR values calculated on the basis of six text segments, each comprising 1,000 words. Using this approach, this paper seeks to explain the lexical richness found in the three books and to examine the relationship between text genre

and the level of lexical variation. It is hoped that these empirical findings will enrich the framework of contemporary linguistic theory and serve as a reference for corpus-based research on Indonesian-language texts.

In line with these research objectives, the theoretical foundations regarding lexical richness and the Mean Segmental Type-Token Ratio (MSTTR) need to be explained as the conceptual basis of this study. Lexical richness refers to the degree of diversity in the vocabulary used within a text. This concept is important for identifying the linguistic characteristics, readability, and communicative functions of a text (Zheng, 2025). Systemic functional linguistics posits that lexical choice is always influenced by the field, tenor and mode of communication (Halliday&Hasan, 1985). Thus, lexical variation is not merely a matter of linguistic aesthetics, but is also closely linked to the communicative purpose and social function of the text.

Lexical choice is also linked to genre. Personal narrative texts tend to use a more varied vocabulary to evoke experiences and emotions, whilst scientific and technical texts rely more on consistent terminology to ensure precision of meaning (Biber et al., 1998). Recent research shows that lexical diversity can distinguish between genre characteristics and the quality of writing (Akbar & Jarvis, 2023; Zhang, 2022). Furthermore, there is a positive correlation between morphological awareness (including derivation and affixes) and a person's lexical richness (Cheng & Li Yin, 2025; Levin & Hovav, 2007). In the context of the Indonesian language, 'memper-kan' is one of the affixed plural word patterns in Indonesian (Solihah et al., 2021). Qualitatively, lexical richness can also be assessed in terms of morphological variation, variation in cohesion markers, and the contextual use of words. In contrast to qualitative measures, the quantitative measure that can be used to assess lexical richness is the MSTTR. The MSTTR is calculated by dividing the participant's description into segments of 40 words and averaging the TTR value for each segment.

METHOD

This study employs a mixed-methods approach that combines quantitative and qualitative analysis (Indrawan & Jalilah, 2021). The quantitative approach was carried out using an explanatory technique based on the MSTTR instrument. The qualitative approach, meanwhile, was carried out using corpus analysis techniques, including frequency analysis using the concordance plot instrument, morphological analysis, and contextual analysis using the Leipzig-Corpora corpus (Ermanto et al., 2023). This approach was chosen so that

the results of the numerical analysis could be enriched with a more in-depth linguistic explanation.

The research data consisted of three Indonesian-language books of different genres, namely *Visi Anak Pasar* by E. Aminudin Aziz (autobiography), *Yang Fana Adalah Waktu* by Sapardi Djoko Damono (novel), and *Tata Bahasa Baku Bahasa Indonesia* by the Language Development and Guidance Agency (reference). From each book, six text segments were selected, each comprising 1,000 words, so that the total data analysed from each book comprised 6,000 words. Thus, the total amount of data was 18,000 words. The number of six segments was determined based on the minimum equivalent data available across all three books. For each 1,000-word segment, the TTR value was calculated by comparing the number of unique words with the total number of words. The TTR value for each segment was then averaged to obtain the MSTTR value for each book. This approach was chosen to minimise text-length bias and to enable more reliable comparisons across genres.

Data were collected using documentation and recording techniques (Husnullail & Jailani, 2024). The text from each book was converted into digital form, then stripped of non-textual elements, such as page numbers, footnotes and visual elements not forming part of the main content. The text was then tokenised into individual words and divided into six segments of equal size, each comprising 1,000 tokens. Segmentation was carried out sequentially in accordance with the order of the text to preserve the original discourse structure.

Data analysis was carried out using explanatory methods (Sari et al., 2022). MSTTR figures are not treated as stand-alone data, but are analysed in relation to the characteristics of the text, the depth of the text, and the communicative purpose of each text genre. Furthermore, data validity is ensured by normalising the number of tokens in each segment so that each text is analysed on an equivalent scale. All data is processed using the same tokenisation procedure to ensure consistency in the calculations. Procedural validation was carried out by gradually rechecking the tokenisation results to ensure there were no errors in token segmentation, punctuation removal, or the handling of specific lexical forms. During the morphological categorisation process, an inter-rater agreement test was carried out on a portion of the data to ensure consistency in the classification of word forms. Furthermore, all stages of the analysis were systematically documented, including the software, software version, analysis parameters, and the scripts or procedures used, so that the research process could be audited and replicated. Methodological triangulation was carried out by

combining the results of quantitative analysis using the Mean Segmental Type–Token Ratio (MSTTR) with qualitative findings regarding vocabulary usage, thereby making the interpretation of the research results more comprehensive, valid and consistent.

RESULTS AND DISCUSSION

The results of the analysis show that the autobiography *Visi Anak Pasar* has the highest average MSTTR value, at 51.5 per cent. This value was derived from six consecutive segments with MSTTR values of 51 per cent, 56 per cent, 52 per cent, 49 per cent, 51 per cent and 50 per cent. The variation between chapters was relatively small, indicating consistency in vocabulary usage throughout the text.

The novel *Yang Fana Adalah Waktu* recorded an average MSTTR value of 45.9 per cent. The MSTTR values per segment varied quite significantly, namely 43 per cent, 37 per cent, 48 per cent, 52 per cent, 50 per cent and 45 per cent. These fluctuations reflect the stylistic and structural dynamics within the literary text. Meanwhile, the book *Tata Bahasa Baku Bahasa Indonesia* has the lowest average MSTTR value, at 39 per cent, with segment-by-segment figures of 45 per cent, 41 per cent, 31 per cent, 39 per cent, 43 per cent and 36 per cent. These values indicate a higher tendency towards lexical repetition compared to the other two books.

No.	Jumlah Kata	Buku Rujukan Ilmiah	Novel	Biografi
		<i>Tata Bahasa Baku Bahasa Indonesia</i>	<i>Yang Fana Adalah Waktu</i>	<i>Visi Anak Pasar</i>
1.	1.000	45	43	51
2.	1.000	41	37	56
3.	1.000	31	48	52
4.	1.000	39	52	49
5.	1.000	43	50	51
6.	1.000	36	45	50
7.	6.000	39	46	51,5

Tabel 1. Table of MSTTR Segmentation Scores

The high MSTTR score in the book *Visi Anak Pasar* indicates that the biographical text possesses the greatest lexical richness compared to the two books in other genres. The MSTTR values, ranging from 49 per cent to 56 per cent across these six segments, do not merely reflect the author's linguistic ability, but primarily reflect the breadth and diversity of the subject's life experiences as recorded. Biographies chronicle the subject's life journey across various social spheres, time periods, professional roles and interpersonal relationships. This diversity of experience demands the use of varied vocabulary to adequately describe events, emotions, social contexts and personal reflections. Thus, the

high MSTTR in these biographical books can be understood as a linguistic trace of the subject's rich and complex life experiences. The high MSTTR reflects the breadth of the world and experiences presented in the text (Halliday&Hasan, 1985; Ma & Chai, 2022).

The linguistic traces of Aminudin's life in the book *Visi Anak Pasar* (Aziz, 2025) are evident not only from the average number of unique words per thousand words—which indicates a rich lexicon—but also from the distinctive vocabulary found in this book. There are several unique lexical items, in the form of words and phrases, that have great potential to be included in the 'KBBI' as new entries, namely the expressions 'kegurubesaran', 'paruh waktu' and 'bingkai takdir'. Some terms, which are colloquial expressions, also have the potential to be included in the KBBI, such as the word 'nguli'. Lexical richness can also be understood through additional dimensions, such as depth of meaning and lexical sophistication (Cheng & Li Yin, 2025; Levin & Hovav, 2007; Zhao, 2024). This is also evident in this book, which features several distinctive expressions that characterise experiences, such as '*mereka mau*', '*saya arungi*', '*mereka omongkan*' and '*saya akrabi*'.

To confirm the value of MSTTR segmentation in the book *Visi Anak Pasar*, one may examine the microlinguistic aspect, namely the diversity of morphological processes within its vocabulary. In this regard, it is found that although the book belongs to the narrative genre, it possesses a varied and complex vocabulary. Several words exhibit diverse affixation processes, such as the -itas suffix in the word 'kollegialitas', the 'ber-i' prefix as in 'berkorespondensi', the 'me-isasi' prefix as in 'memobilisasi', and the 'me-frasa-an' prefix as in 'ketidakmengertian'.

In a broader context, it can be observed that the use of words such as 'kollegialitas' and 'berkorespondensi' has been recorded in the Leipzig Corpus. The selected vocabulary found in the book *Visi Anak Pasar*, when examined there, reveals the following collocation patterns.

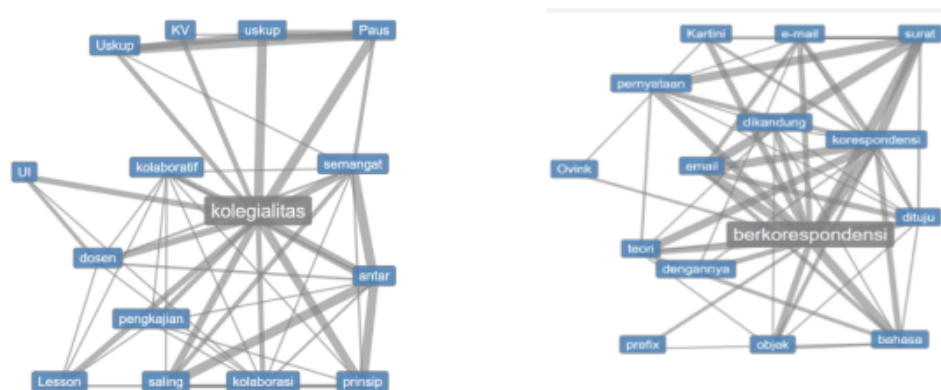


Figure 1. Word Collocations in the Leipzig Corpora

Although the book *Tata Bahasa Baku Bahasa Indonesia* is replete with vocabulary featuring complex morphological processes, there are some forms that do not appear in it, such as the 'ber-i' and 'ber-isasi' forms. The novel *Yang Fana Adalah Waktu* tends to use simple vocabulary with affixes commonly found in everyday life. The table below supports the MSTTR values as outlined above.

Buku	Afiks Lain	Afiks					
		<i>ber-i</i>	<i>per-i</i>	<i>ber-isasi</i>	<i>-itas</i>	<i>di-isasi</i>	
	<i>Visi Anak Pasar</i>	ada	ada	ada	ada	ada	ada
	<i>Tata Bahasa Baku Bahasa Indonesia</i>	ada	tidak ada	ada	tidak ada	ada	ada
	<i>Yang Fana Adalah Waktu</i>	ada	tidak ada	tidak ada	tidak ada	tidak ada	tidak

Table 2. Tabulation of Affix Variability

It should be emphasised that the lexical richness of a biography is relative, not an absolute characteristic inherent to the genre. The degree of lexical variation depends heavily on the author's ability, the subject of the biography, the subject's social and educational background, the sphere of life in which they were active, and the extent of their interaction with various social and cultural spheres. These factors will result in different lexical profiles across biographies. In other words, the MSTTR in biographies reflects not only the text genre but also the spectrum of experiences and the world of meaning lived by the biographical subject. Thus, it is possible that other biographies may not exhibit the same level of lexical richness as found in *Visi Anak Pasar*.

The novel *Yang Fana Adalah Waktu* (Damono, 2018) ranks second in terms of lexical richness. This result indicates that literary works are often associated with linguistic creativity. Consequently, lexical richness tends to be high. In this context, the second-place ranking in the MSTTR segmentation is due to the presence of lexical repetition as a form of aesthetic expression or beauty. The beauty of language means that literary works frequently utilise lexical repetition as a means of building cohesion, rhythm and symbolism. Consequently, in this instance, the MSTTR score for the novel *Yang Fana Adalah Waktu* is not higher than that of the biography *Visi Anak Pasar*. The fluctuations in the MSTTR value between segments in this novel, ranging from 37% to 52%, reflect the dynamics of the

narrative focus, shifting between reflective and descriptive passages rich in vocabulary and more repetitive dialogue-driven sections as presented in the novel (Akbar & Jarvis, 2023; Zhang, 2022).

In terms of morphology—or the diversity of word-forming affixes—this novel ranks last. Literary works that depict people’s everyday lives often employ vocabulary closely associated with daily life, meaning that morphological complexity in the vocabulary is rarely evident. Lexical diversity in a novel is also determined by stylistic requirements and the reader’s engagement with the story. The relationship between lexical richness and the quality of writing is contextual; therefore, the text must be read in accordance with its communicative purpose. (Krzemińska-Adamek, 2016).

Vocabulary distribution is necessary to confirm the lexical richness of each book. The following figure shows a scatter plot analysing the morphological process of the affix ‘memper-kan’, the cohesion marker ‘adapun’, and the choice of words conveying ‘rasa’. The morphological aspect of ‘memper-kan’ was purposively selected as an example of a complex affix. This affix commonly appears in academic texts to indicate the complexity of the subject matter as well. The cohesion marker ‘adapun’ is used to indicate the interconnection of concepts and ideas, and commonly appears across various genres. As for ‘rasa’, it represents the word choice typically found in narrative texts. The scatter plots for each book are shown below.

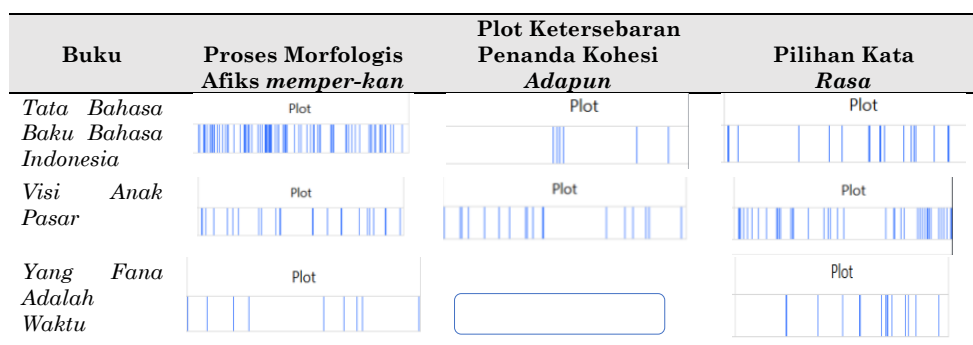


Table 3. Distribution Plot

The lowest MSTTR value was found in the scientific reference book *Tata Bahasa Baku Bahasa Indonesia* (Standard Indonesian Grammar), with the lexical variation in this book falling within the range of 31%–45%. This finding is interesting because scientific texts are often assumed to be difficult and complex. However, the complexity of scientific texts is not always synonymous with high lexical diversity. There are two possible explanations: firstly,

the high number of tokens, which is inevitably linked to the book's length; and secondly, complex sentence structures, even when formed using a relatively limited vocabulary.

On the other hand, this low MSTTR value may also indicate a direct consequence of the text's scientific and prescriptive function, whereby consistency in terminology and the repetition of technical terms are, in theory, prerequisites for clarity and precision of meaning. A comparison of these three books shows that the MSTTR segmentation value cannot be interpreted normatively as 'the higher, the better'. MSTTR values do not always align with text genre, but are closely related to the function of the text in language use. Scientific genres do not necessarily possess a high level of lexical richness. Upon closer examination, this lack of lexical diversity actually consists of specialised terminology.

Autobiographies, as reflective narrative texts in this context, exhibit the highest lexical variation as they represent a wide range of experiences; novels, as aesthetic and imaginative texts, occupy a middle ground as they combine variation and repetition as strategies for aesthetic effect; whilst grammar books, as expository-scientific texts, exhibit the lowest lexical variation as they emphasise terminological consistency. This finding reinforces the view that lexical richness cannot be assessed normatively on the basis of genre alone without considering morphological aspects, the distribution of word choices, their prevalence in a broad corpus, and also the strength of the subject matter.

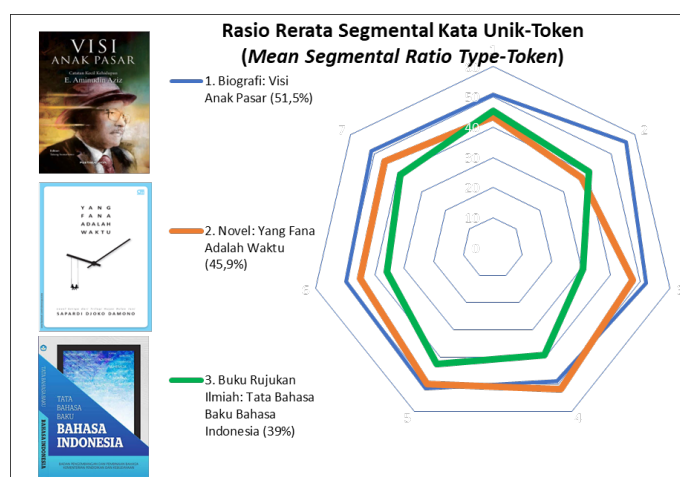


Figure 2. Unique-to-Token Ratio Network Diagram

In the context of language learning, these findings indicate that each genre offers different potential for vocabulary enrichment. Autobiographical texts have the potential to broaden one's vocabulary repertoire through the diversity of experiences and social contexts they present. Novels have the potential to develop sensitivity to meaning, stylistic features, and the use of words within an aesthetic context. Reference books, meanwhile, serve to

strengthen mastery of conceptual terminology and the accuracy of academic language use. Thus, the relationship between lexical richness and language proficiency is not direct, but rather depends on the type of competence developed through the reader's interaction with each genre (Hanušková et al., 2025; Sung et al., 2024). Based on these findings, it can be concluded that lexical richness cannot be interpreted solely on the basis of text genre, but must be understood in relation to the communicative function served by each type of text. Therefore, MSTTR measurements need to be interpreted in conjunction with morphological variation and lexical distribution in order to provide a more comprehensive picture of the characteristics of vocabulary usage.

CONCLUSION

Based on the analysis carried out, it can be concluded that the MSTTR (Mean Segmental Type-Token Ratio) clearly reflects the linguistic differences between the three books, which represent different genres. The autobiography exhibits the highest lexical richness as it represents the protagonist's extensive and varied life experiences. However, in other autobiographical books, this level of richness is relative and depends on the character and life journey of the protagonist, meaning it will not be the same for all autobiographies. Novels display a moderate level of lexical variation with stylistic strategies characteristic of literature, whilst grammar books show the lowest level of variation due to the demands of terminological consistency, which in this context must also be linked to their readability.

In practical terms, this study demonstrates that autobiographies hold great potential as reading material for school pupils and university students. Through the richness of vocabulary and the diversity of experiences depicted, readers not only acquire new vocabulary but also broaden their social, cultural and reflective horizons. By reading biographies, school pupils and university students can 'borrow the life experiences of others', which is then reflected in improved language proficiency and intellectual maturity. Therefore, biographical books deserve to be positioned as a strategic learning resource in language and literacy education.

This study is limited by the small sample size of just three books; consequently, the findings cannot yet be generalised to the entire range of Indonesian-language text genres. Future research could expand the corpus by including more books, authors and genres to provide a more representative picture of the lexical richness of the Indonesian language.

ACKNOWLEDGMENTS (OPTIONAL)

The author would like to express his gratitude to E. Aminudin Aziz, author of *Visi Anak Pasar*; Sapardi Djoko Damono, author of *Yang Fana Adalah Waktu*; and the editorial team of the *Tata Bahasa Baku Bahasa Indonesia* (Standard Indonesian Grammar), 4th Edition, at the Language Development and Guidance Agency. Thanks are also extended to the reviewers of this article.

DECLARATIONS

Ethical considerations

This study adhered to the principles of research ethics throughout the research process. The data analyzed consisted of learners' written texts that were anonymized prior to analysis to protect participants' privacy and confidentiality. The data were used exclusively for academic purposes, and no personally identifiable information was disclosed in this study. All research procedures complied with the ethical standards of the affiliated institution and relevant research regulations.

Use of artificial intelligence (AI)

The authors declare that the generative artificial intelligence (AI) tool [quillbot] was used exclusively for language editing and/or grammatical improvement. The use of AI did not influence the scientific content, study design, data analysis, data interpretation, results, or conclusions of the manuscript. Full responsibility for the content remains with the authors.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

REFERENCE

- Akbary, M., & Jarvis, S. (2023). Lexical diversity as a predictor of genre in TV shows. *Digital Scholarship in the Humanities*, 38(3), 921–936. <https://doi.org/10.1093/llc/fqad004>
- Al Pansori, M. J., Rismawati, L., & Pransisca, M. A. (2025). *Pengantar linguistik edukatif: Teori dan aplikasi dalam pembelajaran bahasa*. CV Eureka Media Aksara.
- Aziz, E. A. (2025). *Visi anak pasar*. Pustaka Jaya.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

- Cheng, X., & Li Yin. (2025). Morphological awareness and vocabulary knowledge in monolingual and bilingual learners: a meta-analysis. *Language Teaching Reseach. Sagepub*.
- Covington, M. A., & McFall, J. D. (2010). Cutting the gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. <https://doi.org/10.1080/09296171003643098>
- Damono, S. D. (2018). *Yang fana adalah waktu*. Gramedia Pustaka Utama.
- Dosi, I., & Anthi, D. (2025). A case study on the narrative skills of a high-functioning autism spectrum disorders child. *International Journal of Research Studies in Education*. <https://doi.org/DOI:10.5861/ijrse.2025.25012>
- Ermanto, S. P., Ardi, H., & Juita, N. (2023). *Linguistik korpus: Aplikasi digital untuk kajian dan pembelajaran humaniora*. PT. Raja Grafindo Persada-Rajawali Pers.
- Halliday&Hasan. (1985). *Language, context, and text: aspects of language in a social-semiotic perspective*. Deakin University Press.
- Hanušková, M., Kubát, M., & Nogolová, M. (2025). Lexical diversity of Czech L2 texts at different proficiency levels. *Glottometrics*, 59, 47–58. https://doi.org/10.53482/2025_59_427
- Hardiansyah, A., Setiawan, A. A., & Sandi, M. (2025). Pengembangan aplikasi segmentasi citra medis untuk deteksi tumor otak dengan akurasi tinggi. *International Journal of Informatics*, 1(1).
- Hardinanto, E. (2025). Potret ideologi perempuan dalam karya sastra feminis: Perspektif analisis wacana kritis. *Unisma Press*.
- Haryani, H., Judijanto, L., Haryono, P., Susandi, N. K. A., Prasetyani, N. Y., & Fatubun, R. R. (2025). *Linguistik terapan*. PT. Sonpedia Publishing Indonesia.
- Husnullail, M., & Jailani, M. S. (2024). Teknik pemeriksaan keabsahan data dalam riset ilmiah. *Jurnal Genta Mulia*, 15(2), 70–78.
- Indrawan, D., & Jalilah, S. R. (2021). Metode kombinasi/campuran bentuk integrasi dalam penelitian. *Jurnal studi guru dan pembelajaran*, 4(3), 735–739. <https://doi.org/10.30605/jsgp.4.3.2021.1452>
- Krzemińska-Adamek, M. (2016). Lexis in writing: investigating the relationship between lexical richness and the quality of advanced learners' texts. In *Second Language Learning and Teaching* (pp. 185–196). Springer science and business media Deutschland GmbH. https://doi.org/10.1007/978-3-319-30373-4_12
- Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 18(2), 154–170. <https://doi.org/10.1080/15434303.2020.1844205>
- Levin, B., & Hovav, M. R. (2007). Morphology and lexical semantics. In *The Handbook of Morphology* (pp. 248–271). wiley. <https://doi.org/10.1002/9781405166348.ch12>
- Ma, Q., & Chai, X. (2022). The relationship between lexical richness and the quality of csl learners' oral narratives. In D. M., G. Y., & H. J. (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): 13250 LNAI* (pp. 95–105). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-06547-7_7
- Mastutik, N. A. P., & Taufiq, M. . A. (2021). Analisis pemilihan kosakata dalam buku ajar perspektif Rusydi Ahmad Thu'aimah. *Al-Ittijah: Jurnal Keilmuan dan Kependidikan Bahasa Arab*, 12(2), 119–132. <http://dx.doi.org/10.32678/al-ittijah.v13i2.5042>
- McCarthy, P. M., & Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2). <https://doi.org/10.3758/BRM.42.2.381>

- Mu'in, F. (2025). *Linguistik terapan: teori dan praktik*. CV Eureka Media Aksara.
- Muldasari, M. (2025). *Kesesuaian Tingkat Keterbacaan Teks Pada Buku Pembelajaran Basa Ugi Pasuloi= The Appropriateness of Text Readability Level in Basa Ugi Pasuloi Learning Book*. Universitas Hasanuddin.
- Purba, N. A., & Sidebang, R. (2024). *Konsep dasar bahasa dan sastra Indonesia*. EDUPEDIA Publisher.
- Sapiah, S., Judijanto, L., Bunga, J., Sholihah, R. A., Krisifu, A., Kocu, S., & Rochma, A. F. (2025). *Pengantar linguistik*. PT. Green Pustaka Indonesia.
- Sari, M., Rachman, H., Astuti, N. J., Afgani, M. W., & Abdullah, R. (2022). Explanatory survey dalam metode penelitian deskriptif kuantitatif. *Metode*, 1(1).
<https://doi.org/10.47709/jpsk.v3i01.1953>
- Savila, A. G. (2025). *Peringkasan teks cerita Islam berbahasa Indonesia menggunakan long short term memory (LSTM)*. Skripsi: Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Solihah, A., Rasyid, Y., & Gomo, T. (2021). *Pengayaan kosakata berbasis tambatan leksikal*. Program Studi Doktor Linguistik Terapan Sekolah Pascasarjana UNJ.
- Suhardijanto, T., & Kusumawardani, R. P. (2021). Kajian awal densitas leksikal untuk pengembangan pelabelan otomatis kelas kata Bahasa Jawa. *Kongres Internasional Masyarakat Linguistik Indonesia*, 404–409.
- Suhendra, S., & Patriani, A. (2021). Analisis penggunaan kohesi leksikal pada teks cerita pendek siswa kelas XI SMA Negeri 4 Bogor. *Triangulasi: Jurnal Pendidikan Kebahasaan, Kesastraan, Dan Pembelajaran*, 1(1), 14–22.
- Sung, H., Cho, S., & Kyle, K. (2024). An empirical evaluation of lexical diversity indices in l2 korean writing assessment. *Language Assessment Quarterly*, 21(2), 159–180.
<https://doi.org/10.1080/15434303.2024.2311728>
- Zhang, X. (2022). The relationship between lexical use and L2 writing quality: A case of two genres. *International Journal of Applied Linguistics (United Kingdom)*, 32(3), 371–396.
<https://doi.org/10.1111/ijal.12420>
- Zhao, M. (2024). Sememe richness: a supplementary indicator of lexical richness in Chinese second-language writing. *Chinese as a Second Language Research*, 13(2), 183–213.
<https://doi.org/10.1515/caslar-2024-2002>
- Zheng, W. (2025). Lexical richness viewed through lexical diversity, density, and sophistication. *Digital Scholarship in the Humanities*, 40(2), 692–708.
<https://doi.org/10.1093/llc/fqaf023>