ENGLISH LANGUAGE AND ARTS EDUCATION JOURNAL

Volume 2 No 2
Year 2025
E-ISSN 3046-6024

Analyzing Student Writing: A Corpus Linguistics Study

Ridha Hasnul Ulya[1*], Wuri Syaputri[2], Siti Fitriati[3], Nursyamsi[4], Aditya Rachman[5], Zulfikarni[6], Dewi Anggraini[7]

[1,5,6,7] Universitas Negeri Padang, Indonesia
[2]Universitas Andalas, Indonesia
[3]Universitas Muhammadiyah Pringsewu, Indonesia
[4]National Research and Innovation Agency (BRIN), Indonesia

[*]Corresponding author email: ridhasnulya@fbs.unp.ac.id

Abstract
The aim of this study was to examine the written compositions of 40 fifth-semester students at Universitas Negeri Padang. The objective was to determine their strengths and areas that need improvement. The study utilised a manual analysis technique, investigating factors such as word frequency, sentence length, grammatical structures, contextual usage, and collocations. This thorough method yielded an in-depth comprehension of the pupils' writing proficiencies. The findings unveiled notable disparities in writing habits among various academic fields. Scientific writing typically employs a straightforward and factual style, characterised by shorter phrases. In contrast, humanities assignments sometimes involve longer sentences that include comprehensive explanations and critical assessments. Students with a higher level of proficiency showed a wider range of vocabulary and better correctness in grammar. In contrast, students with a lower level of proficiency used simpler words and made more frequent grammatical mistakes, such as inconsistent tenses and problems with subject-verb agreement. The majority of students demonstrated accurate usage of common collocations, while there were some instances of improper usage, suggesting a need for a better grasp of these terms. The results also highlighted a dependency on simple conjunctions and a deficiency of more sophisticated transitional words, which affected the writing's coherence and refinement. Based on these findings, the study recommends implementing focused teaching tactics to target specific areas of difficulty. To increase students' writing skills, it is crucial to provide them with targeted grammar teaching, engaging vocabulary tasks, and many opportunities to practice collocations. Moreover, providing discipline-specific writing education can assist students in modifying their writing styles to suit various academic circumstances. Future studies should explore the integration of automated analytic methods and broaden the scope to encompass a wider range of student samples and multimodal data in order to gain a more thorough insight into students'.
Key Words: *Collocations, Grammatical Accuracy, Lexical Diversity*

Submit date: 1 Oct 2025 accept date: 3 Nov 2025 publish date: 5 Nov 2025
Correspondence author: Ridha Hasnul Ulya, Universitas Negeri Padang, Indonesia. E-mail: ridhasnulya@fbs.unp.ac.id
Doi:

Introduction

Examining student writing has great potential to enhance our comprehension of language utilisation, acquisition, and growth (Tressyalina et al., 2024; Yu & Liu, 2021; Zhang & Hyland, 2022). Within educational settings, writing serves as more than just a method of evaluation; it is a crucial instrument for both learning and cognitive advancement. Through student writing analysis, educators and researchers can gain valuable knowledge about students' understanding and use of language rules, structures, and conventions (Aull, 2020; Ferdiansyah et al., 2023; Tamrin et al., 2024). Comprehending this concept is vital for formulating efficient instructional approaches that improve literacy and communication abilities.

Writing is an intricate cognitive procedure that encompasses various interconnected elements, including vocabulary, grammar, syntax, and discourse (Atmazakı et al., 2023; Rachman et al., n.d.; Sudharshana, 2021). Each of these components represents distinct facets of language proficiency and growth. For example, the way students use vocabulary in their writing can demonstrate the extent and complexity of their knowledge of words, whereas the grammatical structures they employ indicate their comprehension of language rules. Likewise, the arrangement and consistency of texts demonstrate students' ability to organize their ideas in a logical and convincing manner.

Corpus linguistics provides a strong and methodical method for analysing written language (Egbert et al., 2020; Ermanto et al., 2024; Meyer, 2023). A corpus refers to a substantial and organised collection of texts that can be computationally processed to reveal patterns and trends that may not be easily discernible through physical examination. When it comes to student writing, creating a corpus entails gathering a wide variety of written texts from students in various grades, disciplines, and levels of ability. This corpus functions as a comprehensive data repository for analysing several language characteristics, including word frequency, collocations (frequent word combinations), and syntactic structures (Eguchi & Kyle, 2023; He & Ang, 2023).

A key advantage of corpus-based analysis is its ability to uncover prevalent usage patterns and stylistic inclinations (Ardi et al., 2018; De Sutter & Lefer, 2020; Shimal, 2022). For instance, it has the capability to recognise the words or phrases that pupils frequently employ, analyse their sentence structure, and detect the usual grammatical mistakes they make (Alghazo & Alshraideh, 2020; Indriyani et al., 2023; Syaputri, 2016). This information can help educators identify areas where pupils may have difficulties and tailor their training accordingly. For example, if a thorough examination of a collection of written or spoken texts reveals that students frequently make errors in the use of specific grammatical structures, educators might create targeted instructional sessions to tackle these problems. Moreover, examining a collection of student writing might provide insight into the progression of writing abilities over a period of time. Researchers can analyze writings from various grade levels to detect shifts in language usage and patterns of development (Biber et al., 2021; Episiasi et al., 2022; Wijayanti et al., 2024). Curriculum design and instructional practices can be

informed and adjusted to match students' changing linguistic abilities by adopting a longitudinal perspective.

To summarize, the use of corpus linguistics in the analysis of student writing provides a powerful means of understanding language usage and progression. Through the construction and examination of a collection of student writings, researchers might discover significant data regarding prevalent usage patterns, stylistic inclinations, and challenging locations. These findings can provide valuable insights for educational practices and facilitate the creation of efficient solutions to enhance students' literacy and communication skills.

Method

The methodology employed in this study entails a meticulous and hands-on process of gathering and examining written samples from 40 fifth-semester students at Universitas Negeri Padang. The data gathering procedure starts by choosing written projects from these students, making sure to include a diverse selection of texts that cover different subjects and topics related to their study. Prioritizing ethical issues is of utmost importance; each student is provided with information regarding the research objectives and the intended use of their written assignments before collecting the samples. Explicit consent is obtained from all students involved, ensuring adherence to ethical guidelines and honoring their entitlement to privacy and secrecy (Alnajjar, PhD & Abou Hashish, PhD, 2021; Bos, 2020).

After the data has been gathered, the corpus is created. The written assignments are carefully categorized into a repository, which can exist in either physical form. Every text is systematically classified according to pertinent metadata, which includes student characteristics such as age and gender, assignment particulars such as subject and date, and any other relevant information that can assist in the analysis (Aljohani & Muslih, 2022).

The qualitative analysis explores the specific ways in which language is used in different contexts. This task entails analyzing the circumstances in which ordinary words and phrases are utilized by studying written materials to comprehend how students utilize certain language and structures in various situations. An analysis is conducted to examine how pupils use conjunctions to establish connections between ideas or adverbs to modify actions in order to understand the functional responsibilities of language elements. The analysis also identifies recurring themes and stylistic traits, highlighting the same topics, narrative styles, and rhetorical methods used by the students.

Collocation analysis entails a manual search and documentation of frequent word pairings within texts. This task requires carefully examining the assignments to identify and document frequently occurring word combinations, as well as subsequently analyzing the settings in which these collocations occur in order to gain a deeper understanding of their usage. For instance, the research may examine the utilization of specific adjective-noun combinations to depict ideas across various disciplines.

An essential aspect of the investigation involves comparing different levels. Cross-level analysis entails examining writing patterns among fifth-

semester students to uncover developmental trends and variations. This involves categorizing the assignments according to specific indications of skill and analyzing the linguistic characteristics exhibited in each category, such as differences in vocabulary usage, sentence structure, and grammatical correctness. Contextual analysis investigates the variations in writing patterns across different subjects and types of assignments. This stage entails analyzing the linguistic characteristics utilized in essays across various disciplines to ascertain subject-specific language usage. It also requires assessing whether students apply distinct stylistic and structural elements when completing narrative versus expository writing assignments.

Result

The study manually analyzed written samples from 40 fifth-semester students at Universitas Negeri Padang. The analysis covered many characteristics of linguistic usage, such as word frequency, sentence length, grammatical structures, contextual usage, and collocations. The findings offer a thorough comprehension of students' writing tendencies, areas of proficiency, and areas requiring enhancement. During the quantitative analysis, the process of manually tallying word frequencies and measuring sentence durations uncovered clear and identifiable trends among the subjects examined. Scientific disciplines typically employ a vocabulary that consists primarily of phrases related to research and methods, which reflect the analytical nature of these tasks. The average sentence length in scientific subjects was 15 words, reflecting the preference for concise and factual writing styles in scientific literature. Sentence lengths varied between 10 and 25 words. In contrast, the vocabulary used in the humanities and social sciences was more diverse, encompassing phrases that were associated with broader concepts and theories. The average sentence length was 20 words, which corresponds to the expository and narrative styles commonly found in these fields. Sentence lengths varied between 15 and 30 words. The results are succinctly presented in the table provided below:

Table 1. Frequent Words

Category	Top 10 Frequent Words	Average Sentence Length	Range of Sentence Lengths
Scientific Subjects	research, data, results, analysis, methods, experiment, hypothesis, significant, variables, measurement	15 words	10-25 words
Humanities and Social Sciences	culture, society, theory, impact, study, perspective, discourse, context, historical, interpretation	20 words	15-30 words

Scientific disciplines mostly employ terminology that is specialised to research and methods, which reflects the analytical nature of these tasks. The students' emphasis on empirical and methodological rigour was evident through their frequent use of terms like "research," "data," "results," "analysis," and "methods." In scientific writing, the typical sentence length was shorter, with an

average of approximately 15 words per sentence. This statement's brevity demonstrates the preference for a brief and factual writing style in scientific literature, where clarity and precision are of utmost importance. The sentence lengths ranged from 10 to 25 words, indicating a preference for concise, straightforward sentences that effectively communicate information without excessive complexity.

The humanities and social sciences, in contrast, employ a wider range of terminology associated with more expansive notions and theories. Frequently employed terms such as "culture," "society," "theory," "impact," and "study" are indicative of the more interpretive and discursive characteristics of these fields. The mean sentence length was greater, at around 20 words, consistent with the explanatory and narrative styles prevalent in these disciplines. The sentence lengths varied between 15 and 30 words, indicating a writing style that is more detailed and descriptive. This technique generally includes more intricate phrase structures to express sophisticated arguments and assessments.

The grammatical analysis uncovered a high frequency of complicated sentences, particularly in argumentative and analytical articles. The intricacy is crucial for formulating and upholding intricate arguments, although it also gives rise to prevalent grammatical mistakes. The errors observed included the improper application of verb tenses, with pupils frequently intermingling past and present tenses in a single paragraph, as well as issues with subject-verb agreement and the erroneous use of articles. For instance, pupils often neglected to include articles when they were required, resulting in a disruption of the grammatical precision of their sentences. These errors were widespread in many disciplines, suggesting the necessity for targeted grammar education to tackle these particular concerns.

During the qualitative research, a thorough investigation of the contextual utilisation of words revealed that students consistently employed academic terminology correctly within their specific circumstances. Nevertheless, there was an excessive dependence on specific words and patterns, which suggests a restricted range of vocabulary. The frequent use of phrases such as "in this research" and "the purpose of this study" indicates that students should improve their academic vocabulary and acquire other methods to convey similar concepts. The investigation of the functional responsibilities of linguistic features revealed that students adeptly utilised conjunctions and transitional phrases to effectively establish connections between concepts and uphold coherence in their writing. Nevertheless, there was a conspicuous excessive utilisation of fundamental conjunctions such as "and" and "but" and a deficient utilisation of more sophisticated connectors like "therefore," "consequently," and "moreover," which could augment the logical progression and refinement of their arguments.

The tasks frequently addressed recurring themes, such as the influence of technology on education, environmental concerns, and cultural analysis. The texts frequently exhibited a formal academic tone, which is suitable for assignments at the university level. Nevertheless, certain assignments had excessive wordiness and repetition, which had a negative impact on the clarity and brevity of the compositions. This implies that although students are skilled at adopting a

professional manner of speaking, they may encounter difficulties in preserving clarity and brevity, especially when addressing intricate topics.

We examined the texts during the collocation analysis to manually identify frequent word combinations, also known as collocations. "Data analysis," "research findings," "literature review," and "case study" comprise the prevalent collocations. These word combinations were very common in assignments that involved research methodologies and academic writing, suggesting a strong knowledge of typical academic expressions. Nevertheless, the study indicated that although students typically employed these collocations accurately, there were occurrences of improper usage, implying that certain students could benefit from supplementary guidance on the utilisation of collocations to guarantee meticulous and precise academic writing. Frequent word pairings, or collocations, were identified manually within the texts. The most common collocations included:

Table 2. Collocation

Collocation	Frequency	Typical Context
Data analysis	35	Used in research and methodology sections
Research findings	32	Common in conclusion and discussion sections
Literature review	30	Found in introductory chapters
Case study	28	Used in descriptions of specific examples or experiments
Theoretical framework	25	Often in background and literature sections

The occurrence of these word combinations was especially common in assignments that focused on research methods and academic writing, suggesting that students demonstrated a strong understanding of typical academic expressions such as "data analysis," "research findings," "literature review," and "case study." Acquaintance with academic vocabulary is essential for generating research-oriented texts that are clear and precise. Nevertheless, a more thorough analysis of these collocations revealed a combination of positive and negative aspects. Despite the majority of students demonstrating a solid understanding of acceptable collocation usage in their works, there were some significant cases where students used collocations incorrectly or without a clear comprehension of their meaning. These misuses indicate that students may not completely understand the subtle differences or appropriate situations for using these common phrases. This suggests that extra education on collocation use is necessary to improve their academic writing skills.

When analysing the writing patterns of students at different proficiency levels in the fifth-semester cohort, we saw several variances that revealed the varying levels of writing competency among students. Students with a higher level of competence demonstrated a greater mastery of vocabulary and grammatical structures, utilizing intricate phrases with fewer grammatical mistakes. The students demonstrated a superior understanding of academic collocations and effectively incorporated them into their writing, resulting in more cohesive and refined compositions. In contrast, pupils with lower proficiency demonstrated restricted lexical variety, frequently relying on less complex language and sentence constructions. Grammatical problems, such as erroneous tense usage and subject-verb agreement issues, plagued their essays. Furthermore,

they struggled to use collocations correctly and frequently misused basic academic words. These findings emphasise the patterns of growth in writing skills among students, indicating that focused instructional assistance could enhance the writing abilities of students with lesser competence and enhance their comprehension of academic collocations. This analysis, conducted across different proficiency levels, highlights the significance of customised instructional methods that cater to the individual requirements of students. These techniques aim to narrow the gap and enhance the overall quality of academic writing across the group. When comparing writing patterns across different proficiency levels within the fifth-semester cohort, several differences emerged:

Table 3. Writing Patterns Across Different Proficiency Levels

Proficiency Level	Lexical Diversity	Sentence Complexity	Common Errors
Higher Proficiency	High	Complex	Few grammatical errors, advanced vocabulary
Lower Proficiency	Low	Simple	Frequent grammatical errors, limited vocabulary

Higher proficiency students demonstrated greater lexical diversity, more complex sentence structures, and fewer grammatical errors. They were adept at using a wide range of vocabulary and constructing intricate sentences that effectively conveyed nuanced ideas. Their writing displayed a strong command of grammar, with minimal errors, showcasing their ability to adhere to academic writing conventions. In contrast, lower proficiency students relied heavily on simpler vocabulary and sentence structures, often producing more straightforward and less sophisticated texts. Their writing was marked by more frequent grammatical mistakes, including incorrect tense usage, subject-verb agreement issues, and misplacement of articles. This stark contrast in writing abilities highlights the developmental trends and variations among the students, underscoring the importance of targeted support. Providing lower proficiency students with focused instruction on grammar, vocabulary enhancement, and sentence construction could significantly improve their writing skills, helping them achieve a level of proficiency closer to that of their higher-performing peers. Furthermore, the contextual analysis revealed additional variations in writing patterns across different subjects, illustrating how students adapt their writing styles to meet the specific demands of various academic disciplines. The analysis also revealed variations in writing patterns across different subjects:

Table 4. Variations in Writing Patterns Across Different Subjects

Subject	Writing Style	Common Features
Scientific Writing	Concise, factual	Frequent use of data, results, methods
Humanities Writing	Elaborate, explanatory	Detailed explanations, critical analysis, theoretical contexts

Scientific writing tasks usually consist of brief and concentrated presentations of factual data, sometimes featuring fewer sentences and a straightforward, unbiased style. This style exemplifies the characteristics of

scientific discourse, which places emphasis on clarity and precision in order to successfully convey research findings and techniques. On the other hand, coursework in the humanities involved more detailed explanations and critical evaluations, requiring students to use longer, more intricate words and adopt a more storytelling approach. The disparity suggests that students possess the skill to modify their writing styles to fulfil the particular requirements of various academic fields, demonstrating their capacity to transition between succinct, objective reporting in scientific settings and more comprehensive, analytical writing in the humanities. The ability to adapt is critical for achieving academic success in a variety of disciplines, indicating the need for discipline-specific writing education to further refine these skills.

The analysis revealed several prevalent areas of difficulty among students, including grammatical precision, lexical variety, and the proper application of collocations. The pupils frequently made grammatical errors, such as inconsistent use of tenses and incorrect subject-verb agreement, which affected the overall grammatical structure of their sentences. These findings indicate the need for targeted grammar training to tackle these common problems. Furthermore, the lack of variety in the vocabulary used by numerous students in their writing suggests that they heavily depend on a small set of words. This emphasizes the need to implement techniques to broaden their range of word options. The improper use of collocations highlights the importance of instructing pupils on the accurate application of common academic expressions.

These findings suggest the use of specific instructional tactics. To enhance students' language skills, it is crucial to design a structured approach to teaching grammar that explicitly targets prevalent mistakes, such as incorrect tense usage and subject-verb agreement. Students can improve by using tailored exercises and individualised feedback. Engaging in vocabulary improvement activities is essential for students to expand their lexical range. These activities involve exploring synonyms and participating in context-based vocabulary exercises. In addition, it is essential to incorporate collocation practice into the curriculum, utilising activities specifically designed to enhance students' comprehension and accurate utilisation of prevalent academic collocations in their written work.

The findings also indicate the need for curriculum development that corresponds to the pupils' language requirements and developmental phases. This includes activities and exercises that explicitly focus on the identified areas of difficulty. For instance, by including regular writing workshops and peer review sessions, students can have chances to practice and improve their writing abilities through repetitive feedback and cooperative learning. By prioritizing these specific teaching methods and improvements to the curriculum, educators can effectively assist students in cultivating the strong writing abilities that are essential for achieving academic excellence in various subjects.

Discussion

The results of this study are consistent with and build upon prior research in the area of student writing and corpus linguistics, offering a detailed perspective on the advantages and disadvantages of existing student writing

methods. Nevertheless, they also possess certain constraints and domains that require additional investigation. This debate will conduct a thorough examination of the advantages and disadvantages of our findings in comparison to other studies, thereby offering a full assessment of their importance and consequences.

An important benefit of our study is that it agrees with earlier research, specifically Jucker (2021) and Lee (2021) findings, which showed that using corpora to look at grammatical features in different genres works. Our study's findings demonstrate that writing styles in different disciplines are closely associated with disciplinary standards. Specifically, scientific writing tends to be brief and focused on presenting facts, whereas humanities writing tends to be more detailed and interpretative. This discovery emphasizes the need to acknowledge and instruct on these distinctions to assist students in adjusting their writing to different academic environments.

Furthermore, our research has practical implications for teaching methods that are in line with Ma et al. (2022, 2024) focus on the advantages of using corpus-based approaches in language instruction. Our findings offer clear guidance for focused instructional tactics by identifying precise areas of difficulty for pupils, including grammatical accuracy, lexical diversity, and collocation usage. These suggestions can assist educators in customising their instructional approaches to target individual student requirements, potentially resulting in enhanced writing results. The actual implementation of our discoveries is a notable advantage, since it directly converts research into actionable educational methods.

In addition, our thorough analysis, which evaluates several linguistic elements such as word frequency, sentence length, grammatical structures, contextual usage, and collocations, offers a comprehensive and nuanced comprehension of student writing patterns. This comprehensive method provides a detailed perspective on students' strengths and areas requiring more assistance, which is essential for creating successful educational interventions. The comprehensive analysis of writing patterns enables a more tailored approach to instruction, thereby effectively addressing the specific needs of each learner.

Nevertheless, our study is limited, especially in terms of its extent. Concentrating solely on fifth-semester students from one university may not accurately reflect the characteristics of larger student populations. Prior studies, such as the analysis conducted by Yoon & Römer (2020) on the Michigan Corpus of Upper-level Student Papers (MICUSP), included a broader spectrum of student levels and institutions, resulting in findings that are more applicable across different contexts. The limited scope of our study may restrict the generalizability of its findings to different contexts, indicating the necessity for future research to incorporate a more heterogeneous sample of students from varying educational backgrounds and institutions.

Another constraint is the dependence on manual analytic techniques, which may increase human fallibility and diminish the reliability of the results. Our study did not have access to advanced corpus linguistics software, like the British Academic Written English (BAWE) corpus analysis by (Nesi & Gardner, 2018), which offers precise and complete analysis. The absence of these

techniques in our study could potentially impact the accuracy and comprehensiveness of the linguistic patterns observed. This suggests the necessity of incorporating automated analytic methods in future research to improve precision and dependability.

In addition, the singular emphasis on written tasks disregards alternative modes of student communication, such as oral language. Gkoumas et al. (2021) and McLean et al. (2020) highlighted the significance of examining many modalities in order to have a comprehensive understanding of language ability. Our study limits its scope to written texts, and incorporating spoken data could provide a more comprehensive view of students' language usage and skills. This constraint highlights the necessity for future research to use multimodal analysis in order to obtain a more comprehensive understanding of students' language capabilities.

Upon comparing our findings with other research, we have identified both positive aspects and limitations. A careful manual analysis of our data is similar to what Gkoumas et al. (2021) found about specific genres, but it doesn't look at as much data as the larger MICUSP and BAWE corpora do. The pedagogical advice we provide aligns with del Mar Sánchez Ramos (2020) research, which emphasises the importance of using corpus-based approaches in teaching. However, the manual nature of our research underscores the need for automated techniques to enhance accuracy and reliability.

Conclusion

The study's manual analysis of written samples from 40 fifth-semester students at Universitas Negeri Padang has yielded useful insights into different elements of linguistic usage, such as the frequency of words, the length of sentences, grammatical structures, contextual usage, and collocations. The results unveiled clear trends in students' written work across several academic fields, emphasising both their areas of proficiency and aspects that require enhancement. Students with a higher level of competency had a wider range of vocabulary and better correctness in grammar, whereas students with a lower level of proficiency relied on simpler words and made more frequent grammatical mistakes.

The research highlights the significance of focused teaching methods to tackle certain areas in which students face difficulties, such as maintaining grammatical precision, enhancing lexical diversity, and employing appropriate collocations. Providing targeted grammar instruction, enhancing vocabulary, and practicing collocations can effectively strengthen student writing skills. Moreover, the study indicates the necessity of providing discipline-specific writing education to better assist students in adjusting their writing styles to fit the requirements of various academic disciplines.

Upon comparing our findings with prior studies, we identified both positive aspects and constraints. Our study confirms previous research on the advantages of using corpus-based methods and emphasises the significance of acknowledging genre-specific writing standards. However, it also underscores the drawbacks of manual analysis and the necessity for more comprehensive and automated approaches. The limited scope of solely examining one university

restricts the applicability of the results, indicating that future investigations should incorporate a more heterogeneous group of students from different educational backgrounds and institutions.

Overall, this study offers a thorough comprehension of the writing patterns of students at Universitas Negeri Padang, providing valuable insights for improving language instruction and learning. Future studies should focus on overcoming the restrictions by integrating automated technologies, broadening the analysis scope, and adding multimodal data to comprehensively assess students' linguistic ability. These procedures will improve our understanding of student writing and facilitate the creation of more efficient instructional approaches to enhance literacy and communication abilities in educational environments.

References

Alghazo, K. M., & Alshraideh, M. K. (2020). Grammatical Errors Found in English Writing: A Study from Al-Hussein Bin Talal University. *International Education Studies*, *13*(9), 1–9.

Aljohani, T., & Muslih, M. (2022). *Learner Profiling: Demographics Identification Based on NLP, Machine Learning, and MOOCs Metadata* [PhD Thesis, Durham University]. Alnajjar, PhD, H. A., & Abou Hashish, PhD, E. A. (2021). Academic Ethical Awareness and Moral Sensitivity of Undergraduate Nursing Students: Assessment and Influencing Factors. *SAGE Open Nursing*, *7*, 237796082110267. https://doi.org/10.1177/23779608211026715

Ardi, H., Nababan, M. R., & SANTOSA, R. (2018). Characters' Politeness Strategies in Giving Command: Should Translators Keep Them? *3L: Language, Linguistics, Literature*, *24*(2).

Atmazakı, A., Ramadhan, S., & Indrıyanı, V. (2023). Dialogic-Interactive Media In Online Learning: Effectiveness In Speaking Skills. *Turkish Online Journal Of Distance Education*, *24*(4), 95–111.

Aull, L. L. (2020). *How students write: A linguistic analysis*.

Biber, D., Gray, B., Staples, S., & Egbert, J. (2021). Investigating Grammatical Complexity in L2 English Writing Research*: Linguistic Description versus Predictive Measurement. In *The Register-Functional Approach to Grammatical Complexity* (pp. 432–457). Routledge.

Bos, J. (2020). *Research ethics for students in the social sciences*. Springer Nature.

De Sutter, G., & Lefer, M.-A. (2020). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, *28*(1), 1–23. https://doi.org/10.1080/0907676X.2019.1611891

del Mar Sánchez Ramos, M. (2020). Teaching English for Medical Translation: A Corpus-Based Approach. *Iranian Journal of Language Teaching Research*, *8*(2), 25–40.

Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.

Eguchi, M., & Kyle, K. (2023). L2 collocation profiles and their relationship with vocabulary proficiency: A learner corpus approach. *Journal of Second Language Writing*, *60*, 100975.

Episiasi, E., Syaputri, W., Suramto, S., & Kasriyati, D. (2022). Lexical and Grammatical Cohesion in the Undergraduate Studentsâ€™ Abstracts. *Linguistic, English Education and Art (LEEA) Journal*, *5*(2), 143–152.

Ermanto, E., Triwira Dhika JR, V., Ardi, H., & Juita, N. (2024). The Use of Confixes Per-/-an and PeN-/-an in the Novel Laskar Pelangi: A Corpus Linguistic Analysis. *International Journal of Society, Culture & Language*, *12*(1), 239–255.

Ferdiansyah, H. M., Rizal, F., & Ulya, R. H. (2023). Problem-Based Blended Learning Models in Vocational Education: A Developmental Research. *International Journal of Information and Education Technology*, *13*(12).

Gkoumas, D., Li, Q., Lioma, C., Yu, Y., & Song, D. (2021). What makes the difference? An empirical comparison of fusion strategies for multimodal language analysis. *Information Fusion*, *66*, 184–197.

He, M., & Ang, L. H. (2023). Profiling a microeconomics noun collocation list: A corpus-based approach. *Southern African Linguistics and Applied Language Studies*, *41*(2), 191–209. https://doi.org/10.2989/16073614.2022.2117708

Indriyani, V., Atmazaki, A., & Ramadhan, S. (2023). Development of Multiliteracy Integrative Learning (MULGRANING) Model in Language Learning. *Education & Science/Egitim ve Bilim*, *48*(215).

Jucker, A. H. (2021). Features of orality in the language of fiction: A corpus-based investigation. *Language and Literature: International Journal of Stylistics*, *30*(4), 341–360. https://doi.org/10.1177/09639470211047751

Lee, J. (2021). Using corpus analysis to extend experimental research: Genre effects in L2 writing. *System*, *100*, 102563.

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning*, *35*(9), 2731–2760. https://doi.org/10.1080/09588221.2021.1895225

Ma, Q., Yuan, R. (Eric), Cheung, L. M. E., & Yang, J. (2024). Teacher paths for developing corpus-based language pedagogy: A case study. *Computer Assisted Language Learning*, *37*(3), 461–492. https://doi.org/10.1080/09588221.2022.2040537

McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, *37*(3), 389–411. https://doi.org/10.1177/0265532219898380

Meyer, C. F. (2023). *English corpus linguistics: An introduction*. Cambridge University Press.

Nesi, H., & Gardner, S. (2018). The BAWE corpus and genre families classification of assessed student writing. *Assessing Writing*, *38*, 51–55.

Rachman, A., Syaputri, W., Kasriyati, D., Episiasi, E., Mulyanti, D., & Retnawati, S. (n.d.). Language Proficiency and Academic Success in University: A Comprehensive Study. *Community Practitioner*, *21*(2), 202–212.

Shimal, M. (2022). *A Corpus-Based Stylistic Analysis of Body Language in the Novels of Charles Dickens and Naguib Mahfouz* [PhD Thesis, The University of Texas at San Antonio].

Sudharshana, N. P. (2021). From Cognitive Grammar to Pedagogic Grammar: Macrostrategies for Designing Form-Focused Tasks. In N. P. Sudharshana & L. Mukhopadhyay (Eds.), *Task-Based Language Teaching and Assessment* (pp. 163–181). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-4226-5_9

Syaputri, W. (2016). Peer Evaluation to Teach Semantics. *Jurnal Smart*, *2*(2).

Tamrin, M., Lestari, R. W., Ananda, A., Syaputri, W., & Alfurqan, A. (2024). The Impact Of The Jigsaw Learning Model On The Educational Outcomes And Citizenship Of Grade V Students At SD IT Nurul Ikhlas Padang. *Community Practitioner*, *21*(1), 361–378.

Tressyalina, T., Syaputri, W., Ginanjar Sapari, G., Fitriati, S., Afnita, A., Noveria, E., & Liusti, S. A. (2024). Decoding Cultural Signifiers: Semantic Analysis of Bahasa Minangkabau in Everyday Discourse. *Community Practitioner*, *21*(2), 1317–1328.

Wijayanti, R. A., Rahwanda, D., Efendi, Y., & Syaputri, W. (2024). Challenging The Students by Using Video Movie to Improve Students' Vocabulary on Independent Curriculum. *ELT-Lectura*, *11*(1), 13–24.

Yoon, H.-J., & Römer, U. (2020). Quantifying Disciplinary Voices: An Automated Approach to Interactional Metadiscourse in Successful Student Writing. *Written Communication*, *37*(2), 208–244. https://doi.org/10.1177/0741088319898672

Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, *48*, 100525.

Zhang, Z. V., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, 100586.