



Does Quartile Matter? Investigating syntactic complexity of international publication

Laily Martin

Department of English, Faculty of Humanities, Universitas Andalas, Padang, Indonesia

lailymartin@hum.unand.ac.id
Tel *: +6281363288678

Abstract

With the challenge of international publication, this study compares the syntactic complexity of Indonesian scholars' publications. The analysis covers 21 journal articles from two groups, from the journals with quartile and without quartile. Using 14 syntactic complexity measures, the results show that the journal articles with quartile have higher mean scores of syntactic complexity measures than the non-quartiles. However, significant differences only occur in three groups of measurement: Length of Production Unit, Coordination, and Degree of Phrasal Sophistication. The findings may show the performance gap between groups as the syntactical constructions of the journal articles with quartile surpass the non-quartile.

Keywords: EFL writers, syntactic complexity, Indonesia

eISSN: 2398-4287 © 2023. The Authors. Published for AMER & cE-Bs by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers), and cE-Bs (Centre for Environment-Behaviour Studies), College of Built Environment, Universiti Teknologi MARA, Malaysia
DOI: <https://doi.org/10.21834/e-bpj.v8iSI16.5216>

1.0 Introduction

This study compares the syntactic complexity of journal articles written by Indonesian scholars who have successfully published their research internationally. The discussion covers fourteen measures of syntactic complexity, which are grouped into five types: Length of Production Unit, Sentence Complexity, Subordination, Coordination, and Degree of Phrasal Sophistication (Lu, 2010). The idea was based on the scarce publication discussing linguistic aspects of journal articles at the lower level. Previous studies on journal articles within the Indonesian context mainly discuss linguistic aspects at the discourse level: rhetorical structure (Argument pattern (Bermani, Safnil, & Arono, 2017)), style (writing argumentatively (Arsyad, 2018); review and citation styles (Arsyad, Zaim, & Susyla, 2018)), as well as linguistic features and devices to develop the discourse (tense and aspect of citation (Arsyad, Zaim, & Susyla, 2018)). These studies demonstrate how ideas are presented in journal articles; there are lack of records on how more basic writing components, such as syntactic elements, are used in particular writings. One of them, syntactic complexity, can be used to evaluate academic writing since it correlates with writing quality (Kyle & Crossley, 2017; 2018). In this study, the relationship between syntactic complexity and writing quality focuses on the writing performance by the journal ranking, particularly quartile, which is used as an indicator of journal prestige in the Indonesian context of international publication. Even though the relationship between linguistic performance and journal ranking is still debatable (Lu et al., 2019), the topic is still understudied.

Furthermore, with the development of technology used in linguistic studies, syntactic complexity can be analyzed using larger corpora and user-friendly tools, like Syntactic Complexity Analyzer (Lu, 2010), which helps to describe the syntactic complexity of texts. The software contains syntactic complexity measures from previous studies that can be chosen for a specific investigation. The program has been used to study second language learners' works (Johnson, 2017; Yoon, 2017), but the indexes can describe advanced-level writers' syntactic choices (Yin, Gao, & Lu, 2021). To fill the gap of minimum existence of syntactic complexity studies of journal articles, this study aims at comparing the syntactic complexity of the journal articles written by Indonesian scholars, the journals with quartile and without quartile, by seeking the answer to the question "Are there significant differences between the journal articles with quartile and without quartile in terms of the syntactic complexity? If yes, what are those differences?"

eISSN: 2398-4287 © 2023. The Authors. Published for AMER & cE-Bs by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers), and cE-Bs (Centre for Environment-Behaviour Studies), College of Built Environment, Universiti Teknologi MARA, Malaysia
DOI: <https://doi.org/10.21834/e-bpj.v8iSI16.5216>

2.0 Literature Review

2.1 Studies of syntactic complexity in academic writing

Syntactic complexity is "the range and degree of sophistication of syntactic structures that surface in language production" (Ai & Lu, 2013, p.249). By this statement, the investigation of syntactic complexity measures the linguistic performance of texts, spoken or written, focusing on the variety and complexity of syntactic structures used in texts. Some possible causes of differences in a text's syntactic complexity are text topic (Yoon, 2017), proficiency levels (Yoon, 2017), writing development (Drijbooms, Groen, & Verhoeven, 2016), age (Drijbooms, Groen, & Verhoeven, 2016), task types (Nippold et al., 2014; 2015), Text types (Beers & Nagy, 2011; Thorne, Fischer, & Lu, 2012), and writers' background (Ai & Lu, 2013).

In academic writing, syntactic complexity has been studied concerning linguistic performance. It has been claimed to contribute to writing quality, even regarded as one indicator of writing quality (Biber, Gray, & Staples, 2016; Kyle & Crossley, 2017, 2018; Yang, Lu, & Weigle, 2015). In a later study, Biber, Gray, Staples, and Egbert (2020) underlined the necessity of studying students' writing performance and development with linguistically interpretable analyses, such as syntactic differences. In particular to journal articles, the studies on syntactic complexity were related to rhetorical functions and levels of writers. Lu, Casal, and Liu (2020) investigated syntactic complexity and rhetorical functions in the Introduction section of journal articles on social science. The results showed that writers used different syntactic complexity to realize the rhetorical functions of the Introduction section. Another study by Yin, Gao, and Lu (2021) demonstrated the syntactic complexity across different part-genres of journal articles: Abstract, Introduction, Literature Review, Method, Results, Discussion, and Conclusion. They found the different degrees of syntactic performances between the Chinese emergence and expert writers across the part-genres. Then there is one study that relates syntactic complexity and journal ranking, which was conducted by Lu et al. (2019), claiming that syntactic complexity had no relation to the scientific impact of journals.

2.2 Measurement of syntactic complexity

There are two main approaches to measuring syntactic complexity, measuring the length of language production and counting the number of linguistic units within a bigger construction. For example, Nippold (2010) proposes two measures of syntactic complexity: mean length and density. Mean length focuses on the average number of clauses per utterance (MLU), per terminal unit (T-unit, MLTU), or per communication unit (MLCU), and density measures the Clausal Density (CD). Later, Lu developed more detailed and concrete measures of syntactic complexity (2010, 2011); he synthesized fourteen syntactic complexity measures from the studies of previous experts, which are grouped into five syntactic categories (Lu, 2010, pp.478-479), as described in Table 1. He then collaborated with Ai (Ai & Lu, 2013) to utilize these measures in an online tool of syntactic complexity measures (L2SCA) and claimed that the sets are the result of various and prolonged studies on syntactic complexity (Lu, 2017).

Table 1. The measures of syntactic complexity

Measure	Code	Definition
Type 1: Length of Production Unit		
Mean Length of Clause	MLC	#of words/# of clauses
Mean Length of Sentence	MLS	#of words/# of sentences
Mean Length of T-Unit	MLT	#of words/# of T-unit
Type 2: Sentence Complexity		
Clauses per Sentence	C/S	# of clauses/# of sentences
Type 3: Subordination		
Clauses per T-unit	C/T	#of clauses/# of T-unit
Complex T-units per T-unit	CT/T	#of complex T-units/# of T-units
Dependent Clauses per Clause	DC/C	#of dependent clauses/# of clauses
Dependent Clauses per T-unit	DC/T	#of dependent clauses/# of T-units
Type 4: Coordination		
Coordinate Phrases per Clause	CP/C	# of coordinate phrases/ # of clauses
Coordinate Phrases per T-unit	CP/T	# of coordinate phrases/# of T-units
T-units per Sentence	T/S	# of T-units/# of sentences
Type 5: Degree of Phrasal Sophistication		
Complex Nominals per Clause	CN/C	# of complex nominals/# of clauses
Complex Nominals per T-unit	CN/T	# of complex nominals/# of T-units
Verb Phrases per T-unit	VP/T	# of verb phrases/# of T-units

(Source: Lu, 2010)

As seen in the table, there are differences in the system of syntactic complexity measurement. They depend on which linguistic unit in which syntactic structure is used as the measuring core. For instance, the measures of Type 1 are based on word occurrences in a specific syntactic structure that the measures count the number of words that occur in a clause, sentence, or T-unit. Rather than counting word occurrence, the measure of Type 2 counts the number of clauses that occur in a sentence, and Type 3 and 4 count the number of embedding (or subordination) and coordination in a particular syntactic structure. Then type 5 measures the complexity of phrasal structure, taking nominal and verb phrase structures as the indicators of sophistication (Ortega, 2003 in Lu, 2010).

3.0 Methodology

3.1 Procedure

The research is text analysis (Wang, 2020), in which the discussion purely derives from the lexical composition of written texts. The comparative analysis compared the linguistic performance, particularly syntactic complexity, between the journal articles written by Indonesian scholars published internationally in 2019. Some of the articles were published in journals with certain ranks; some others do not have ranks. In this study, the ranking system used as the sample is the quartile system, regarded as an accomplishment standard in the Indonesian higher education context. It means that even though publishing internationally in journals with or without quartile is similarly appreciated, publications in journals with quartile have higher prestige and reward than the non-quartile. Therefore, in the selection process later, the journal articles were grouped into two, publications with quartile (Q) and without quartile (NQ), to see the different syntactic complexity. The sample texts were selected based on the similarity of the field of study, English linguistics and English teaching.

3.2 Texts for Analysis

There were 21 journal articles of international publication identified published in 2019; nine are the Qs and the other twelve are the NQs. These articles were the products of Indonesian scholars across different parts of Indonesia who have successfully represented their universities in international academic communication.

3.3 Data Processing and Analysis

Preceding data collection, all 21 texts of the journal articles were converted into Microsoft Word for more editing purposes leaving only the texts' content. In the editing, there was no grammatical editing to maintain the originality of the syntactic performance of the writers. All figures and table contents were omitted, including the headings and subtitles, to leave only the complete syntactic structures. The conversion was done considering the use of the single mode of the online tool. Using the mode allows user to run the program freely by copy-pasting the text directly to the system without the requirement to register to the system. The conversion was also done in order to be able to process the texts later in the Syntactic Complexity Analyzer (L2SCA) program (<https://aihayang.com/software/l2sca/single/>) for collecting the scores of each syntactic complexity measure (Ai, 2016; Ai & Lu, 2010). This program has been used widely to help describe the syntactic performance in spoken or written texts. The program runs automatic calculations for each syntactic complexity measure and provides a score for each measure. The results were transformed into Microsoft Excel to help organize the two data sets (Q and NQ) and process the data with simple statistical procedures. Descriptive statistical information of mean values was counted from both data sets. A rough comparison was made by using the results to see the behavior of the data sets.

To test the Significant Differences of the mean scores between the two data sets, independent sample t-tests were run. The t-scores revealed the degree of significant differences between the two data groups to show the gap in writing performance. As the investigation of syntactic complexity was done to 14 measures (indicating that there were 14 tests performed in the same dataset simultaneously), the Bonferroni correction was also done to set the new alpha value (α), resulting at 0.004. It derived from the alpha value for each comparison 0.05/14, or 0.004, where 0.05 is the significant level for the complete set of tests, and 14 is the number of the individual test being performed.

4.0 Findings

The measurement results using the Lexical Complexity Analyser online program are described in Table 2. There are two scores stated in the table, the mean scores and the t-test scores. The mean scores are used as the brief view of the results of measuring syntactic complexity, and the t-test scores are used to show the results of testing the significant differences in the measures between the Qs and NQs.

Table 2. Syntactic complexity measurement of the quartile and non-quartile journal articles (Q vs. NQ)

Measure	Code	Mean score		Significance
		Q	NQ	

Type 1: Length of Production Unit					
Mean Length of Clause	MLC	15.395	11.522		+
Mean Length of Sentence	MLS	31.207	20.109		+
Mean Length of T-Unit	MLT	28.334	18.291		+
Type 2: Sentence Complexity					
Clauses per Sentence	C/S	2.047	1.754		-
Type 3: Subordination					
Clauses per T-Unit	C/T	1.858	1.598		-
Complex T-Units per T-Unit	CT/T	0.584	0.422		-
Dependent Clauses per Clause	DC/C	0.456	0.331		-
Dependent Clauses per T-unit	DC/T	0.868	0.572		-
Type 4: Coordination					
Coordinate Phrases per Clause	CP/C	0.615	0.292		+
Coordinate Phrases per T-unit	CP/T	1.117	0.472		+
T-units per Sentence	T/S	1.102	1.112		-
Type 5: Degree of Phrasal Sophistication					
Complex Nominals per Clause	CN/C	2.276	1.482		+
Complex Nominals per T-unit	CN/T	4.194	2.373		+
Verb Phrases per T-unit	VP/T	2.878	2.230		-

In general, the Q group has higher mean scores of syntactic complexity measures (except the Coordination, specifically on the T-unit per Sentence) than the NQ, which indicates that the performance of the writers published in the journals with quartile is higher than the non-quartile. The discussion is as follows:

1. Length of Production Unit

As seen in Table 2, the mean scores of Qs are higher than the NQs; the mean scores of the Mean Length of Clause are slightly higher, but the mean scores of the Mean Length of Sentence and Mean Length of T-unit are significantly higher. It means that the Qs contain longer clauses, sentences, and T-units than the NQs as seen in more words used in discussing the topics.

2. Sentence Complexity

On the ratio of using clause per sentence, the Qs contain a slightly higher number of clauses than the NQs, making the sentence constructions longer and more complex.

3. Subordination

The results shown by the measures of Subordination indicate that the Qs contain more clauses in each syntactic structure than the NQs. It shows the ability to produce longer and more complex syntactic constructions using clause structures.

4. Coordination

Similar to Subordination measures, the mean scores of Coordination measures show slightly higher scores of the Qs than the NQs, except on the measure of T-unit per sentence. They indicate the different abilities in producing coordinate phrases in both Clause and T-unit constructions.

5. Degree of Phrasal Sophistication

Again, the mean scores of the measures under this category show that the Qs contain a significantly higher number of Complex Nominals used in clauses and T-units than the NQs. Complex nominals refer to the complex construction of noun groups. The results showing that the Qs have higher scores than the NQs indicates the different degree of sophistication in constructing nominals in the two text groups. Similarly, the writers of the Qs also use more verb phrases in T-unit than those of the NQs.

However, as the Significant Differences were tested, not all measures resulted positively significant; Significant Differences occur on three measure types: The Length of Production Unit (Mean Length of Clauses, Mean Length of Sentence, Mean Length of T-unit), Coordination (Coordinate Phrases per Clause, Coordinate Phrases per T-unit), and Degree of Phrasal Sophistication (Complex Nominal per clause, Complex Nominal per T-unit).

5.0 Discussion

Different from the research results of Lu et al. (2019), who found that journal ranking of impact factor did not have relationship to the syntactic complexity, the current study shows that the journal quartile has relationship to the syntactic complexity; the articles from the journals with quartile contain different syntactic complexity than the articles from the non-quartile journals. The results of t-tests show Significant Differences between the two groups on the following syntactic complexity measures:

1. The Length of Production Unit

All three measures of The Length of Production unit show Significant Differences between the Qs and NQs. It indicates that the Qs contain significantly longer and more complex constructions of clauses, sentences, or even T-units than the other group (NQ).

2. Coordination

Only two measures of Coordination show Significant Differences between the Qs and NQs; both are in the construction of coordinate phrases (either in clause or in T-unit), indicating that the Qs contain more coordinate phrases in clause and T-unit than the number of T-units per sentence.

3. Degree of Phrasal Sophistication

The two measures of Degree of Phrasal Sophistication also show a significant difference. Both are in the constructions of complex nominal occur in clause or T-unit. It shows that the writers of Qs use more sophisticated phrasal constructions than those wrote in the NQs. However, these two groups of writers similarly produce constructions of verb phrases in T-units.

The different findings of the current study are probably caused by the different fields of the journals; the study investigated journal articles on English linguistics and English teaching, while Lu et al. (2019) worked with articles from Biology and Psychology. Lu et al. (2019) have mentioned in their discussion of related works that the relationship between journal ranking and linguistic performance behaves differently in different fields, perhaps by the influence of the writing styles of each field and journal or by the different ranking systems used as the focus of study.

Another factor of differences in syntactic complexity can be related to the level of writers. The writers who have successfully published in the journals with the ranking perform more ability to produce more various and complex syntactic constructions than the non-ranked journals, which results in the higher scores of syntactic complexity measures. The results may also be related to the standard of writing expected by the journals of different ranking; the higher the journal level, the more rigid of selection criteria of the journals, including the standard of the linguistic performance of the articles. This factor may also differentiate the writers' proficiency levels, as Yin, Gao, and Lu (2021), for example, referred to the category as the emergence and expert writers. The above discussion on the different performances of publication resulted from the measurement of the syntactic complexity does not only show linguistic information but also the potential to adopt the computational system in writing assessment to complete the predictive measurement (Biber et al., 2020).

6.0 Conclusion & Recommendations

The syntactic complexity measurement of 21 journal articles published by Indonesian writers in 2019 reveals that the articles successfully published in the journals with quartile commonly have higher mean scores than those with non-quartile. Even though Significant Differences are only in the three measure groups, the results generally give a picture of the performance gap between the two groups. However, using L2SCA tool that helps identify writing performance, as shown by this sample study, should be regarded as the evaluation of writing performance at the initial stage. Judgment to more detailed and comprehensive identification of linguistic performance for writing quality may go to deeper and more complex linguistic analyses.

For broader purposes, automatic tools may also be applicable in writing courses. They help identify the students' writing performance and the level of proficiency in learning management that can bridge into the expected target performance. Measuring with the tools may also be applicable to identify the L2 writers' performance, either students or experts, to compare the performance of writers from different L2 backgrounds, to identify the difficulties faced by writers of different backgrounds, or even to compare the performance of the L2 writers with the writers of English-speaking backgrounds. Such studies can lead to several efforts to boost the publication quality of L2 writers in the future.

Paper Contribution to Related Field of Study

From the measurement to the syntactic complexity, the current study shows the different linguistic performances between the journal articles published with quartile and with non-quartile. It indicates that high-rank journals are more selective with language quality. The findings may inform on which syntactic complexity measures the non-quartile journals are weak, particularly in the context of non-English expert writers. Perhaps by paying more attention to their weak performance of the syntactic complexity, efforts to support potential authors with a lack confidence in English can be proposed.

References

Ai, H. (2016). Web-based L2 syntactical complexity analyzer [Software]. Available from <http://aihaiyang.com/software/l2sca/>

- Ai, H. & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier and P. Thompson (Ed.), *Automatic Treatment and Analysis of Learner Corpus Data* (pp. 249-264). Amsterdam: John Benjamins.
- Arsyad, S. (2018). Struggling for international publication: The potential rhetorical problems for Indonesian scholars in social sciences and humanities when writing in English. *Proceeding ICLA: Advances in Social Science, Education and Humanities Research*, 301, 469-478.
- Arsyad, S., Zaim, M., & Susyla, D. (2018). Review and citation style in research article introductions: a comparative study between national and international English-medium journals in medical sciences. *Discourse and Interaction*, 28-51. <https://doi.org/10.5817/DI2018-1-28>.
- Bermiani, R. A., Safnil, & Arono. (2017). An analysis of argument structure of research article of English postgraduate program of Bengkulu University published in journal. *Journal of Applied Linguistics and Literature*, 2(2), 47-64. <https://doi.org/10.33369/joall.v2i2.5954>.
- Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639-668.
- Biber, D., Gray, B., & Staples, S., and Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, 46, [100869], <https://doi.org/10.1016/j.jeap.2020.100869>.
- Drijbooms, E., Groen, M. A., & Verhoeven, L. (2016). How executive functions predict development in syntactic complexity of narrative writing in the upper elementary grades. *Read Writ*, 30, 209-231. <https://doi.org/10.1007/s11145-016-9670-8>
- Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing*, 37, 13-38.
- Kyle, K. & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102, 333-349.
- Lu, C., Bu, Y., Dong, X., Wang, J., Larivière, V., Sugimoto, C. R., Paul, L. & Zhang, C. (2019). Analyzing linguistic complexity and scientific impact. *Journal of Infometrics*, 13, 817-829. <https://doi.org/10.1016/j.joi.2019.07.004>.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36-62.
- Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493-511, <https://doi.org/10.1177/0265532217710675>.
- Lu, X. and Ai, H. (2015). Syntactic complexity in college-level English writing: differences among writers with diverse L1 backgrounds. *Journal of second language writing*, pp. 1-12. <http://dx.doi.org/10.1016/j.jslw.2015.06.003>
- Lu, X., Casal, J. E., & Liu, Y. (2020). The rhetorical functions of syntactically complex sentences in social science research article introductions. *Journal of English for Academic Purposes*, 44, 1-16.
- Nippold, M. A. (2010). *Language sampling with adolescents*. San Diego, CA: Plural.
- Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C. & MacKinnona, M. (2015). Critical thinking about fables: examining language production and comprehension in adolescents. *Journal of Speech, Language, and Hearing Research*, 58, 325-335.
- Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P., Kirk, C., Hayward-Mayhew, C., & MacKinnon, M. (2014). Conversational and narrative speaking in adolescents: Examining the use of complex syntax. *Journal of Speech, Language, and Hearing Research*, 57, 876-886. [https://doi.org/10.1044/1092-4388\(2013\)13-0097](https://doi.org/10.1044/1092-4388(2013)13-0097)
- Thorne, S. L., Fischer, I. & Lu, X. (2012). The semiotic ecology and linguistic complexity of an online game world. *ReCALL*, 24(3), pp. 279-301.
- Wang, W. (2020). Text analysis. In H. Rose and J. McKinley (eds.) *The Routledge handbook of research methods in applied linguistics* (pp. 453-463). London: Routledge.
- Yang, W., Lu, X., & Weigle, S. A. (2015). Different topics, different discourse: Relationship among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53-67.
- Yin, S., Gao, Y., & Lu, X. (2021). Syntactic complexity of research article part-genres: differences between emerging and expert international publication writers. *System*, 97, [102427]. <https://doi.org/10.1016/j.system.2020.102427>.
- Yoon, H.J. (2017). Linguistic complexity in L2 writing revisited: issues of topic, proficiency, and construct multidimensionality. *System* 66, pp. 130-141.