

Sentence Studies in Indonesian Research Articles from the History and Law Disciplines

Warsidi^{1*}

¹ Universitas Islam Makassar, Indonesia

* warsidi.dty@uim-makassar.ac.id

Abstract

Studies on sentences have been conducted in many linguists' academic works, but only a few of them investigate this area in research articles. Besides, sentence studies in Indonesian research articles have received a very small portion of academic attention. For this reason, the present study intends to investigate words and sentences, as well as voice forms in Indonesian research articles to provide a more comprehensive understanding to students and novice authors. The data in the present study are 30 Indonesian research articles from the disciplines of History and Law. The analysis used a top-down approach, combining a computer analysis system and manual analysis. An earlier analytical approach was employed and the results indicate that both data sets share similarities and differences. Their similarities and differences are presented in both the result section and the discussion section.

Keywords: *Indonesian History and Law research articles, sentence studies, voice forms, word and sentence numbers*

Introduction

Studies on sentence formation have been conducted worldwide, and this area may become pivotal for academic contexts. For example, studies on this area have been carried out on students' WhatsApp daily status (Saputra et al., 2021), natural language processing (Guo et al., 2019), sentence comprehension (Gonzalez Alonso et al., 2021; Jäger et al., 2017), song lyrics (Christianto, 2018), narrative English textbook (Jaya, 2017), Thai English Foreign Language (EFL) Students (Sermsook et al., 2017), research articles (Deveci, 2019; Helal, 2014; Lu et al., 2020; Warsidi et al., 2023; Wu et al., 2020), and many more. In this review, sentence studies in research articles seem more obvious and frequent than in other fields. The reason may be that research articles are considered academic writing, and thus, many linguists want to deeply investigate to provide meaningful comprehension for their readers.

In research articles, sentence studies were conducted in 30 research articles (RAs) in English as a Lingua Franca (ELF) and 30 RAs in American English (AmE) with a focus on syntactic complexity (Wu et al., 2020). The RAs are from four disciplines: natural science, medicine, agriculture and forestry, social science, humanities, and behavioral science. The findings showed that the ELF group has 181.29 mean sentence lengths, while the RAs in the AmE have 113.71 mean lengths. In addition, the ELF authors prolong their sentences at the phrasal level as opposed to the clausal level and use coordinate nouns and adjectives more frequently than the AmE authors. Consequently, compared to the AmE writers' writing, the

RAs in the ELF group contain longer sentences. They aim to furnish more details regarding the knowledge that is involved. Besides,

Lu et al. (2020) used Swales' (2004) CARS model to study the connection between syntactic complexity and rhetorical devices in 600 Research Article Introductions (RAIs) from the disciplines of Applied linguistics, economics, political science, psychology, sociology, and anthropology. The results showed that the corpora have an average of 856 words per RAI and 25.91 words per sentence. The findings also indicate that complex sentences only arise in a few numbers.

Another study about sentences in RAs was investigated in RAIs written by French and American authors (Helal, 2014). First, concerning the length of texts, the French RAIs' sentences are shorter than the American RAIs as they have 9.4 words per sentence, while the American RAIs have 15.6 words per sentence. Also, the French RAIs tend to avoid long texts and minimize their communication of specific content, while the American RAIs communicate their purpose maximally to clarify all the knowledge involved. In addition, French writers tend to compose their RAIs inductively, while American authors compose their RAIs deductively.

Another grammatical study was conducted by (Deveci, 2019), who analyzed 50 RAs written by Turkish authors using English as Lingua Franca (ELF) and 50 RAs written by Anglo-American authors. The RAs in both groups were published in 26 peer-reviewed and open-access journals (see p. 99). The results show that the corpora have 594,611 words, with an average of 5946.11 words per RA. The AngloAmerican authors write longer sentences than Turkish authors, and the AngloAmerican authors employ 24.88 words per sentence on average, while the Turkish authors employ 23.51 words per sentence. Although the Turkish RAs written in English have short sentences, they employ more complex sentences than simple sentences. Both groups also tend to use complex sentences, more particularly in their discussion and introduction sections.

Khedri and Kritsis (2020) discovered a variation in sentence length of English RAs in different disciplines. In the soft sciences, applied linguistic RAs have 7893 words per RA, and psychology RAs use 7552 words per RA. However, in the hard sciences, environmental engineering RAs have 6844 words per RA, and chemistry RAs employ 7168 words per RA. The shortest RA has 3529 words, and the longest RA employs 9982 words. Therefore, the English RAs in soft sciences tend to employ longer texts than those in the hard sciences. Then, Swales (1990) found that in 1893, RAs had 7,000 words per RA. Between the 1970s and 1980s, the number of words decreased to 5,000 words per RA. After the 1980s, however, the average numbers increased to roughly 10,000 words per RA (Swales, 1990).

Another study on sentence writing was undertaken by Kutay (2016), who investigated sentence forms within the RAMs in the medical nursing discipline published in Indian journals. The purpose was to discover whether the authors tend to employ active or passive sentences, whether they tend to use past or present forms, and whether they tend to use relative clauses, modal verbs, or that-clauses. The analysis found that passive sentences are used more frequently (66.18%) than active sentences, which means the authors tend to avoid including their personalities inside texts. They also tend to write sentences in the past tense (80.51%) more than in the present tense and tend to employ relative clauses more than modal verbs and that-clauses, which means that the authors are

inclined to describe and clarify specific terms to help their readers comprehend the arguments.

Arsyad (2014) studied how grammar works in research article abstracts (RAAs). The investigation focused on six parameters that were arranged into two large clusters. Hedge words, attitudinal attitude markers, and self-reference words are components of the second cluster, which is an interactional metadiscursive form (p. 202). Grammar functions related to "voice (active and passive voices), English tense (present and past tenses), and that-complement" comprise the first cluster (p. 202). Most Indonesian RA authors utilize active sentences that are predominantly in the present tense grammatically, with the exception of Move 3 (which discusses the technique), where 50% of the phrases are in the past tense, and Move 4 (which reports the study findings). Hedges are used meta-discursively by most Indonesian RA writers, especially in Move 4.

In short, sentence studies have been analyzed in various fields and indicate important topics for the analysis to provide more sentence comprehension. In this review, although some studies have investigated sentence structures and length, hardly found a study on sentence forms in research articles. Missing this weak pivotal gap in the literature has encouraged the present author to conduct analyses of the Indonesian RAs from the disciplines of History and Law because these two selected disciplines have never been investigated to date. For this intention, the author formulated the following research questions:

1. How many word and sentence numbers are within RAs in these two selected disciplines? Do these two data sets have different word and sentence numbers? If so, to what extent?
2. What voice forms do they employ in the Introduction, Method, Results, and Discussion (IMRD) sections? Do these two data sets have different voice forms? If so, to what extent?

Method

Data Sets

In this study, thirty Indonesian RAs in all were used as data sets, fifteen in each of the two disciplines—law and history. Because this study did not only examine one section of RAs but also analyzed all other major sections, such as the introduction, method, result, and discussion, it was determined that using 15 RAs from each discipline as the corpus of the study would be sufficient for analysis. All these corpora were RAs published in Science of Technology Index (SINTA 2) journals from the disciplines of History and Law.

Data Analysis

To analyze word numbers, each RA was detected using a computer system in each section and their appearances were noted and tabulated in the table form. Then, to analyze sentence numbers, the author reads manually and identifies each sentence whether it is active or passive sentences.

To analyze the voice form, a framework from Mustika (2017) was used to investigate whether the sentences were more prominent in passive or active voice. As presented in the theoretical framework, this framework was used because it clearly defines both passive and

active voices in the Indonesian language. Such clarity is necessary to analyze the data of the present study, as without clarity, the reliability of results could be questioned. Active voice means that the subject of the sentence is the doer of an action (Mustika, 2017). In contrast, the passive voice means that the subject receives an action rather than doing an action. In other words, the subject is the target that gets an effect of the action. These criteria were used to analyze the voice form to discover whether the authors tend to employ passive or active voice in their RA sections and communicative moves. An understanding of the voice form can determine whether the RA authors admit or hide their agency in the RA sections and communicative moves.

In this regard, active voice means that any sentence or clause is structured using the construction *subject + verb + [object]*; in other words, the subject does the action [*verb*]. In contrast, passive voice is any clause or sentence in which the subject does not conduct an action but receives the action. For example, *Tugas itu dikerjakan oleh saudara perempuanku* (The assignment is done by my sister). In this context, the subject *tugas itu* (the assignment) does not do the action, but it receives the action from the doer (agent), that is, 'my sister'. Thus, the passive voice is in contrast to the active voice.

To analyze the voice forms in the RA sections, every sentence was read using a top-down approach to determine whether the sentence was in active or passive voice. Then, for the voice forms in the communicative moves, the focus was on identifying whether the sentence in the moves and steps is in active or passive voice. Thus, the procedures for analyzing the syntactic level and the voice form were alike, as the focus was on analyzing every single sentence using the top-down approach in both their RA sections and communicative moves. The number of voice form appearances was then counted and reported in the table.

Results

Words and Sentence Numbers

After conducting analysis, this study revealed that History research articles have more word numbers than those of Law research articles. Besides, History research articles also have more sentence numbers than those of Law Research articles. More details about word and sentence numbers are presented in Table 1.

Table 1. Total Word and Sentence Numbers within two data sets

Category names	History RAs	Law RAs
Total word numbers	93610	74924
Average word numbers per RA	6240.67	4994.93
Total sentences	3553	2485
Average sentence numbers per RA	236.87	165.67
Word numbers per sentence	25.35	15

As presented in Table 1, the total word numbers for all Indonesian History Research Articles (IHRAs) is 93610 words, with an average of 6240.67 words per RA. This word number is distributed as 3553 sentences, with an average of 236.87 sentences per RA. Each sentence contains 25.35 words on average (these numbers exclude the abstract and conclusion sections). However, the total word numbers of the (Indonesian Law Research Articles (ILRAs) are 74924 words, with an average of 4994.93 words per RA. The authors distribute these

total words into 2485 sentences, with an average of 165.67 sentences per RA. Thus, each sentence contains 15 words on average. These findings indicate that IHRAs have longer texts and contain more information than those in ILRAs.

Voice Forms in Both Data Sets

After analyzing the data, the analysis revealed that the majority of sentences of clauses in the IHRAs are in active voices. For more details, Table 2 reports the summary of the data analysis.

Table 2. Summary Analysis of the Voice Forms in the IHRAs

Sections in the IHRAs	Voice forms		Total clause numbers
	Active voice	Passive voice	
Introduction	585	229	814
Method	118	55	173
Result and Discussion	2071	958	3029
Total clauses in all sections	2774	1242	4016
Percentages	69.07%	30.93%	100%
Average per RA	184.93	82.8	267.73

As presented in Table 2, the active voice is more dominant than the passive voice in all sections, including the introduction, method, results, and discussion sections. The total number of active voices is 2774 (69.07%) clauses, while only 1242 (30.93%) clauses have passive voice. This finding indicates that active voice is conventional in realizing RA sections within the IHRAs because it appears in more than 60% of the total clauses.

Like in IHRAs, the analysis of the voice form in the ILRAs also indicated active voices are more dominant than passive voices. More details of data analysis results are reported in Table 3.

Table 3. Summary Analysis of the Voice Form of the ILRAs

Sections in the ILRAs	Voice forms		Total clause numbers
	Active voices	Passive voices	
Introduction	380	179	559
Method	42	23	65
Result and Discussion	1715	741	2456
Total clauses in all sections	2137	943	3080
Percentages	69.38%	30.62%	100%
Average per RA	142.47	62.87	205.33

As shown in Table 3, most clauses found in the data analysis are in active voice, with 2137 appearances (69.38% of the total clauses). In contrast, the passive voice only appears in 943 clauses (30.62% of the total clauses). As in the IHRA data, the passive form without the 'di-' prefix is not found in the ILRA data either. This finding means that the active voice is used twice as often as the passive voice. Therefore, the active voice is conventional in the ILRAs because it appears in more than 60% of the total clauses. On the other hand, the passive voice is optional in the ILRAs as it appears in less than 60% of the total clauses. More details about the voice forms found in the data analyses are exemplified below.

Active Voice

Active voices appear most both data sets. Some examples of active voice found in the data analyses are:

Example 1: *Untuk selanjutnya para aktivis lingkungan **bukan hanya membatasi** diri pada lingkungan, tetapi juga hal-hal lain yang bersifat unik seperti tumbuhan dan binatang liar yang unik, proteksi terhadap makluk langka.* HIS04

[Henceforth, environmental activists **do not only concentrate** (their activities) on the environment, but also on other things that are unique, such as unique wild plants and animals and protection against rare creatures.]

Example 2: *Pengusaha-pengusaha yang melirik bisnis sinema **berlomba-lomba mendirikan** perusahaan film di berbagai kota.* HIS05

[Entrepreneurs who glance at the cinema business **compete** to establish film companies in various cities.]

Example 3: *Mereka **menjalankan** misi penyelamatan terhadap warga Desa Mapu yang ditahan oleh pasukan kartel internasional pimpinan Panglima Timur (Aqi Alexa).* HIS07

[They **carried out** a rescue mission against Mapu Village residents who were detained by international cartel forces led by the East Commander (Aqi Alexa).]

To indicate whether a clause is in active voice, the author pays attention to the subject and verb. The bold phrases above are verb phrases in the active form. When the subject does the action, it indicates an active sentence. In Example 8.15, the subject does an action, and the verb is also active, which means that the sentence is in the active voice. As seen in all the examples above, therefore, all the bolded verbs are in active form.

Passive Voice

Passive voice only appears in small portion of the total clauses. Some examples of the passive clauses found in the data analyses are:

Example 4: ***Pelaksanaan zakat harus diawasi oleh penguasa, dilakukan oleh petugas dan dipungut** dari orang yang wajib mengeluarkan untuk diberikan kepada yang berhak menerimanya.* LW02

[The implementation of zakat **must be supervised by the authorities, carried out** by officers and **collected** from people who are obliged to pay it, to be distributed to the needy.]

Example 5: ***Keberkahan dimaknai sebagai karunia Tuhan yang terdapat pada diri seseorang atau pada benda tertentu.*** LW02

[**Blessing is interpreted as a gift from God** that is found in a person or a particular object.]

Example 6: *Dengan demikian, **sebuah perjanjian harus dilandasi pada ketaatan pada hukum, kerelaan dan kejelasan poin-poin yang diperjanjikan.*** LW03

[Thus, **an agreement must be based on obedience to the law, willingness, and clarity of the points in the agreements.**]

Example 7: ***Kebolehan poligami dalam Islam hanya dapat dilakukan apabila dapat berbuat adil kepada istri-istri.*** LW04

[Polygamy **is allowed** in Islam only if the husband can be fair to his wives.]

All the bold sentences in the examples above use passive voice because the subjects in all the examples do not conduct an action, but rather they receive an action. In Example 8.38, the subject *pelaksanaan zakat* (the implementation of *zakat* - Islamic religious alm) just

receives an action but does not do the action; therefore, it is in a passive voice. In the other three examples, the subjects are similar, as they do not conduct an action but receive actions. Also, most instances of passive voice in the present findings use a verb followed by the prefix *di*, which means the subject receives an action. For example, *diawasi* (is supervised), *dimaknai* (is interpreted), and *dilandasi* (is based) are verbs with the prefix *di*, which means they are in passive voice.

The Comparison of Both Data Sets

In comparison, both data sets have similarities and differences. However, the author prefers to present the similarity first. The similarity is that both of the data sets tend to use active voices more than passive voices. This result means that authors from both data sets tend to reveal their agency rather than hide it. However, both data sets have differences in word and sentence numbers. In this regard, the analysis indicates that IHRAs contain more word and sentence numbers than those found in ILRAs. Thus, this comparison suggests that they have similarities and differences in realizing sentences in their RAs.

Discussion

As found in the analysis, the IHRAs have more word and sentence numbers than the ILRAs. The IHRAs have 4016 clauses, with an average of 267.73 clauses per RA. On the other hand, the ILRAs only have 3080 clauses, with an average of 205.33 clauses per RA. The findings above indicate that the IHRAs have longer texts than the ILRAs.

However, their voice form in communicating all sections and realizing communicative moves is alike. In communicating their RA sections, the IHRAs tend to write in active voice, as it appears in 69.07% of the entire clauses. Like the IHRAs, the ILRAs also tend to employ active voice (69.38%). This finding indicates that both groups are alike as both mostly employ active voice in all sections. Therefore, both IHRAs and ILRAs tend to reveal their agency in realizing their RA sections.

In realizing RA sections, both the IHRAs and ILRAs are similar as they tend to employ active voice in 69.07% and 69.38% of the total clauses, respectively. This finding means that the authors tend to foreground their agency more than hide it. This finding supports the earlier study of Arsyad (2014), in which it was found that most Indonesian RA abstracts in the disciplines of education, economics and management, and humanities also use active voice. This finding is also similar to most thesis writing, as Yannuar et al. (2014) found that the authors tend to use active voice twice as often as passive voice. Furthermore, it is similar to English RAs (Lightman, 1974; Stoeger, 1980; Tarone et al., 1981, 1998), as most of the verbs in English RAs are in active form. It is also similar to both Iranian and international journals in applied linguistics, as their authors tend to employ active voice more than passive voice (Mohsenzadeh & Ebrahimi, 2017). English RAs in the soft sciences also tend to use active voices (Seoane & Hundt, 2018).

In contrast, the present research findings are different from those found in Seoane and Hundt (2018). These authors found that the authors of hard science RAs tend to employ passive voice. Also, the method section of medical-surgical nursing RAs tends to use passive voices, appearing in 66.18% of the total voices (Kutay, 2016). Sentences within move structures in English civil engineering RAs are also in passive voice (Kanoksilapatham, 2011). All these RAs that tend to employ passive voice are in the hard sciences. Thus, the present

research findings are similar to the findings for English RAs in the soft sciences in that they tend to employ active voice, but they are different from those in the hard sciences. Therefore, the authors in the present research findings tend to admit their agencies more than hide them. In this case, the text is more emphatic because it tends to employ active voice rather than passive voice, and the subject of the text is the obvious executor of the action (Mohsenzadeh & Ebrahimi, 2017).

Conclusion

The present findings showed that Indonesian research articles in the History discipline have more word and sentence numbers than those in the Law discipline. This finding indicates that History RAs have longer texts than those in the Law discipline. However, both of these two data sets share similar voice forms, and both of them tend to employ active voices more than passive ones. This study implies that these two data sets share both similarities and differences.

Unfortunately, this study is only limited to sentence numbers and forms. However, although it has limitations, the findings may contribute to both theoretical and practical applications. Theoretically, these findings may contribute to sentence studies in the literature that active forms are more applicable in writing research papers in all sections. Practically, the findings may be useful for novice authors to adjust their writing when they want to publish an article in Indonesian good journals (such as accredited SINTA 2).

References

- Arsyad, S. (2014). The discourse structure and linguistic features of research article abstracts in English by Indonesian academics. *Online Submission*, 10(2), 191-223.
- Christianto, D. (2018). Syntactic analysis of sentence patterns in John Denver's song lyrics. *Unpublished Thesis, Universitas Sanata Dharma, Indonesia*.
- Deveci, T. (2019). Sentence length in education research articles: A comparison between Anglophone and Turkish authors. *Linguistics Journal*, 13(1), pp, 73-100.
- Gonzalez Alonso, J., Cunnings, I., Fujita, H., Miller, D., & Rothman, J. (2021). Gender attraction in sentence comprehension. *Glossa: a journal of general linguistics*, 6(1). <https://doi.org/10.5334/gjgl.1300>
- Guo, H., Mao, Y., & Zhang, R. (2019). Augmenting data with mixup for sentence classification: An empirical study. *arXiv preprint arXiv:1905.08941*. <https://doi.org/10.48550/arXiv.1905.08941>
- Helal, F. (2014). Genres, styles and discourse communities in global communicative competition: The case of the Franco–American 'AIDS War'(1983–1987). *Discourse Studies*, 16(1), 47-64. <https://doi.org/10.1177/1461445613496352>
- Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, 94, 316-339. <https://doi.org/10.1016/j.jml.2017.01.004>
- Jaya, A. (2017). Sentence patterns of narrative text English textbook in Indonesia. *English Community Journal*, 1(2), 69-74.
- Kanoksilapatham, B. (2011). Civil engineering research article introductions: Textual structure and linguistic characterization. *The Asian ESP Journal*, 7(2), 55-84.

- Khedri, M., & Kritsis, K. (2020). How do we make ourselves heard in the writing of a research article? A study of authorial references in four disciplines. *Australian Journal of Linguistics*, 40(2), 194-217. <https://doi.org/10.1080/07268602.2020.1753011>
- Kutay, U. (2016). A genre analysis of the methodology sections of descriptive medical-surgical nursing articles. *Journal of Computer and Education Research*, 4(7), 65-81. <https://doi.org/10.18009/jcer.09009>
- Lightman, A. P. (1974). Time-dependent accretion disks around compact objects. I-Theory and basic equations. II-Numerical models and instability of inner region. *The Astrophysical Journal*, 194, 419-427. <https://doi.org/10.1086/153259>
- 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, 100832. <https://doi.org/10.1016/j.jeap.2019.100832>
- Mohsenzadeh, A., & Ebrahimi, S. F. (2017). Verbs in applied linguistics research article introductions: Semantic and syntactic analysis. *International Journal of Foreign Language Teaching and Research*, 5(17), 25-33.
- Mustika, N. (2017). Kalimat pasif dalam bahasa Inggris dan bahasa Indonesia: Suatu analisis kontrastif. *Jurnal Kata: Penelitian tentang Ilmu Bahasa dan Sastra*, 1(2), 109-116. <https://doi.org/10.22216/jk.v1i2.1861>
- Saputra, U. R., Maulina, M., Nasrullah, R., & Sakkir, G. (2021). Students' Sentence Errors on WhatsApp Daily Status: A Literature Review. *Celebes Journal of Language Studies*, 23-31. <https://doi.org/10.51629/cjls.v1i1.31>
- Seoane, E., & Hundt, M. (2018). Voice alternation and authorial presence: Variation across disciplinary areas in academic English. *Journal of English Linguistics*, 46(1), 3-22. <https://doi.org/10.1177/0075424217740938>
- Sermsook, K., Liamnimit, J., & Pochakorn, R. (2017). An Analysis of Errors in Written English Sentences: A Case Study of Thai EFL Students. *English Language Teaching*, 10(3), 101-110. <https://doi.org/10.5539/elt.v10n3p101>
- Stoeger, W. R. (1980). Boundary-layer behavior of the flow at the inner edge of black hole accretion disks. *The Astrophysical Journal*, 235, 216-223. <https://doi.org/10.1086/157626>
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1981). On the use of the passive in two astrophysics journal papers. *The ESP Journal*, 1(2), 123-140. [https://doi.org/10.1016/0272-2380\(81\)90004-4](https://doi.org/10.1016/0272-2380(81)90004-4)
- Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1998). On the use of the passive and active voice in astrophysics journal papers: With extensions to other languages and other fields. *English for Specific Purposes*, 17(1), 113-132. [https://doi.org/10.1016/S0889-4906\(97\)00032-X](https://doi.org/10.1016/S0889-4906(97)00032-X)
- Warsidi, Irawan, A. M., Adnan, Z., & Samad, I. A. (2023). Citation studies in English vs. Indonesian research article introductions (RAIs) in the history discipline. *Studies in English Language and Education*, 10(2), 598-613. <https://doi.org/10.24815/siele.v10i2.28343>
- Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*, 43, 100798. <https://doi.org/10.1016/j.jeap.2019.100798>

Yannuar, N., Shitadevi, I. A., Basthomi, Y., & Widiati, U. (2014). Active and passive voice constructions by Indonesian student writers. *Theory & Practice in Language Studies*, 4(7). <https://doi.org/10.4304/tpls.4.7.1400-1408>