JALL (Journal of Applied Linguistics and Literacy)
ISSN 2598 - 8530
English Education Program
Faculty of Teacher Training and Education
Galuh University
Jl. R.E. Martadinata No. 150 Ciamis 46251
jall@unigal.ac.id

https://jurnal.unigal.ac.id/index.php/jall/index
JALL (Journal of Applied Linguistics and Literacy), ISSN 2598-8530, February, Vol. 5 No. 1, 2021

Received	Accepted	Published
December 2020	February 2021	February 2021

LEXICAL BUNDLES OF INDONESIAN AND ENGLISH RESEARCH ARTICLES: FREQUENCY ANALYSIS

Azka Saeful Haq
azka19001@mail.unpad.ac.id
Department of Linguistics, Faculty of Cultural Science Padjajaran University
Rosaria Mita Amalia
Department of Linguistics, Faculty of Cultural Science Padjajaran University
Susi Yuliawati
Department of Linguistics, Faculty of Cultural Science Padjajaran University

ABSTRACT

This study is preliminary research of lexical bundles in the corpus of Indonesian and English research articles that focuses on analysis of frequency and distribution. This study aims to acquire list of common lexical bundles in applied linguistics articles and describes the patterns of bundle use. The most frequent lexical bundles investigated by frequency criteria reflect the common pattern of bundle use in each corpus. Frequency-based approach to multi-word combination enables us to acquire reliable results because of its statistical test in authentic language data. The result shows that the most numerous bundles are 3-word length and surprisingly, 5-word bundles *it can be concluded that* occurs in the top 20 rank in Indonesian corpus. The comparison between corpora reflects that the bundles across text section are identical. Although there are the same bundles used in both corpora, the typical bundles with high score of frequency and range are found to characterize the different group of writers. The distributional patterns show that there is the presence of popular bundles in English and Indonesian writers. The top rank lists emphasize that the common lexical bundle structures are phrase-based in expert level. Practically, this study can play role in English for Academic Purposes (EAP) to recommend prevalent patterns of lexical bundle use in the form of pedagogically useful list of word combination. The findings can also be used for non-native writers or scholars especially Indonesian writers to enrich the use of lexical bundles across sections in language and linguistics field.

Keywords: lexical bundles, corpus-driven approach, frequency analysis,

INTRODUCTION

Research article written by non-native writers is potentially problematical to engage with native writing style. Native-like writing marked by linguistic features in a text is underlain by practice and comprehension that are integrated in language learning. For non-native and novice writers, it is important to improve the quality of their article through learning the native-like writing style in academic genre. In the context of academic community, the writers need to use prevalent academic expression to increase the value of their articles. Learning common writing style can be helpful for high quality research need to be constructed in appropriate writing. Less awareness of the importance of writing style in academic writing becomes a factor that cannot improve the quality of writing.

Research article contains more than selection of academic dictions in lexical aspect. There is the presence of word combination used in specific discipline to reflect particular patterns of use which are crucial for writers. Numerous corpus studies prove the big role of word combinations in research articles that they can be the markers of non-native or native and novice or expert writing through identifying the use of word combination (Breeze, 2013; Chen & Baker, 2010; Cortes, 2013; Hyland, 2008; Hyland & Jiang, 2018; Pan et al., 2016; Salazar, 2014). The existing studies uncover that word combination as linguistic feature in research articles become marker of register, genre, discipline, and academic competence (Salazar, 2014). The studies further recommend that word combination has to become materials in English for Specific Purposes (EAP), not a single academic diction.

The different writing style between native and non-native writers is marked by the common word combination used repeatedly in their writing. Native-like writing competence becomes additional value for an academic work and it can be one of the problems for non-native writers to acquire many chances

in international academic involvement such as publication in reputable international journals (Yuliawati et al., 2020). List of common word combinations that are usually used by native writers in particular discipline can be useful for non-native writers to set their rhetorical style as well as guide in academic writing. Especially for junior scholars, their works need to be recognizable scholarly through using common frequent phrases (Hyland & Jiang, 2018).

The word combination that become the unit of analysis in this study is called in various terminologies namely multi-word unit, n-grams (or specifically bigrams or trigrams), clusters, formulaic language, phraseological sequences, phrasing, chunks, prefabricated patterns and lexical bundles. They as linguistic feature are used frequently by writers and represents the characteristics of academic writing especially research article. Lexical bundles in this study refer to unit of analysis under corpus linguistics as the approach to investigate real language use of a particular discourse community (Biber & Barbieri, 2007).

Significance of lexical bundles studies in academic writing is to provide familiar patterns of use in word combinations for guideline. The linguistic evidences reflected by lexical bundles are useful to be implemented in English for Academic Purposes such as English writing, teaching materials, proficiency test, and syllabus design. The lexical bundles (Biber & Barbieri, 2007; Hyland & Jiang, 2018) represent natural and original language use constructed from communicative experiences in particular discourse community. They are marker to identify characteristics of particular academic writing and to measure conventional patterns of language use.

Previous studies of Indonesian articles (Budiwiyanto & Suhardijanto, 2020a, 2020b; Yuliawati et al., 2020) concern on articles written in Indonesian language and do not deal with the analysis across text sections. The lexical bundles in Indonesian research that is written in English articles need to be explored to acquire enough comprehension in serving our research to a written

English description. The most frequent lexical bundles in Indonesian articles can be compared and contrasted with native English articles to acquire adjustment for further writing. This study aims to investigate native English and Indonesian lexical bundles as an effort to require more native-like writing styles in particular disciplinary communities.

In terms of literary gaps, the specific discipline namely Language and Linguistics subject category or discipline becomes literary gap in this study because the existing studies mostly investigate two or more academic disciplines (Budiwiyanto & Suhardijanto, 2020b; Durrant, 2015; Hyland, 2008; Hyland & Jiang, 2018; Kwary et al., 2017). This study also compares and contrasts four different sections of research article namely introduction, method, results & discussion, and conclusion that become the gaps in investigating Indonesian lexical bundles. Literature review section is not considerably included because of its relatively less presence based on articles that are collected in this study. In order to acquire more efficient analysis, the section of result and discussion are united. The purpose of this effort in this study is to acquire the knowledge of prevalent rhetorical style of different article section in two different group of writers.

This study employs main theory of lexical bundles pioneered by (Biber & Barbieri, 2007) and supported by numerous related studies in word combination or lexical bundles (Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2013; Hyland & Jiang, 2018). The lexical bundles are generated based on frequency-based approach that can handle large language data in electronic form with the help of corpus tool (Nasselhauf, 2005 in (Salazar, 2014).

Lexical bundles theory is under corpus linguistics for it is conducted on the basis of computer supports, mixed method, and large authentic language data. It makes this study empirical in acquiring research goal instead of intuitive language study. Corpus method namely n-grams, tool are used to generate and analyse the bundles automatically.

Frequency becomes the central concept that underpins the analysis of corpus (Baker, 2006) and it is investigated in this study for they can reveal empirically patterns of bundle use in authentic language data. This approach as the most basic statistical test enables us to conduct more quantitative analysis in measuring the presence of lexical bundles. Quantitative data reflect the quantity of the bundle use within different corpus in the numerical form.

The patterns of bundle use found in this study can further be used to improve the writing styles. How to use the bundles in particular discourse community can be learned by individual or the help of instructors in EAP setting. The pedagogical implication of lexical bundles frequency list and composition of bundle that can be implemented in EAP with the specific disciplinary bundles that have been found in studies of lexical bundles (Gavioli, 2005).

METHOD

This study employs mixed method design that involves two forms of data in a single study. This is in line with the study conducted by Farihah & Rachmawati (2020) that employed both qualitative and quantitative analyses in a study. The purpose is to get the comprehensive analysis of data. Quantitative phase in data analyzing is represented by frequency-based approach in the context of identifying the unit of analysis. The approach is aimed to generate frequency amounts of lexical bundles in a list to acquire the most commonly used bundles as well as their structures. Qualitative phase in data analyzing deals with close-reading through investigating context in concordance lines to see the functions of bundle in the text. Both two phases can produce wider understanding to see language use phenomena especially in the use of lexical bundles.

Source of Data

The criteria of intended data for corpus construction are determined based on the purpose of this study namely to investigate lexical bundles in two different domains. General criteria for intended journals as source of data are:

1. Journals concerned on language and linguistics subject category
2. High impact factor journals
3. Using English language in all articles
4. Journal published between 2015 and 2020
5. The open access journal articles

Each criterion contains consideration based on the purpose such as specific area reflected by language and linguistics category and the most numerous citations reflected by high impact factor journals. The articles published between 2015 and 2020 represent updated articles at the time when this study is conducted. The open access articles enable whoever to check easily the selected articles for data validation. After the general criteria are adopted, each corpus needs to be specified in the context of suitability in representing native and non-native or Indonesian academic articles. It reflects the consideration of representativeness in constructing corpus and manifestation of specific purpose in corpus construction. The processes of data selection in compiling research articles are under the criteria and they are conducted manually which mean they are download without any help of software.

The specific criteria for native articles consider the quality that represents reputable international journal articles written by British and American experts. The criterion of native writers is traced through identifying the names of the writers. Articles that are conducted under international collaboration are included if they involve native English writers. Affiliation and tittles that represent a country or specific region can be additional consideration in several cases. The criterion of expert can be found in the articles published in highest

impact factor journals based on Scimago Journal Rank (SJR) website and all journals are quartile 1.

One of the criteria for non-native is Indonesia domain for this study concerns on Indonesian context. Specifically, the journals have to be accredited by Science and technology index or Sinta in its highest score in national scale namely Sinta 1 and 2. All Indonesia journal articles need to be limited in the context of native writers in Indonesian journals. Based on the steps to find journals, Language and Linguistics journals indexed by Sinta (S1) in Indonesia are only four that are eligible and the others are Sinta 2 journals

From the corpus construction process of twenty journals, it is obtained approximately two million tokens. Not all of contents in complete article are included such as literature review section and it decrease automatically the number of tokens. 200 articles are hoped to represent proportional presence of each article from 10 different journals. The 5 years period between 2015 and 2020 is considered to have proportional composition in each of corpora. Corpora of article conclusion become the least number from eight corpora in this present study. The detail of tokens of each corpus are presented in the table below.

Table 1. Corpus Tokens

Article Section	Native English (British & American)	Non-native (Indonesia)
Article Introduction Corpus	137.853	181.086
Article Method Corpus	295.922	117.414
Article Research and Discussion Corpus	723.682	468.436
Article Conclusion Corpus	99.373	52.302
Total of tokens	1.256.830	819.238
Number of Articles	200	200

Corpus Compilation

This study uses corpora that contain research articles in linguistics discipline built from native English (British and American) and non-native or Indonesian journal articles. The process of the two corpora construction is

conducted differently. The differences are in terms of the source and procedure. The detail procedure of each corpora construction will be elaborated further.

In the context of English corpus construction, investigation of journals' profile is conducted for ensuring that every journal is indexed by Scimago Journal Rank (SJR) https://www.scimagojr.com/. The rank that displays impact factor of each journal and go to official journal website is available in SJR website for first dataset. In official homepage of each journal, the *all issue* menu is selected to see holistically the portrait of journals. Article selection is conducted under the criteria that will be explained further and each article is downloaded systematically from the top position to the lower one in journal website.

The non-native or Indonesian corpus is built from different source of electronic scientific database. The second dataset is built upon the investigation in Sinta official website concerned on Sinta 1 category. There are only four journals that are indexed in Sinta 1 and Sinta 2 based on the investigation in query terms. There is no option in Sinta official website to search for the rank in particular subject category, namely language and linguistics in this context. The search column in Sinta 1 https://sinta.ristekbrin.go.id/journals?q=&search=1&sinta=1 search is implemented with the queries namely *language, linguistics, and education* separately but for *education* query must be complemented by *language* or *linguistics* queries. After all of the articles are downloaded, they are grouped in different folders for further converting process.

In the context of representativeness, article downloading process is done per a journal. Each journal which represents various linguistic fields such as language education, translation, discourse, language and computer, and micro linguistics has equal proportion in each corpus. Every journal with its proportional articles is placed in corpus from the last volume in 2020 to the oldest one in 2015.

Published articles are downloaded per volume started from the most updated issues in 2020 to issues in 2015. Each article with *pdf* format is converted to *docx* firstly to clean irrelevant information mostly related to publication. Unintended information such as journal volume description in header or footer is removed including the authors' name and affiliation. References in each article are also deleted for they are not considered as the contents of articles. Compatible format for corpus tool namely plain text format or *.txt* is adopted after all of the texts are cleaned and ready to analyse.

Table 2. Corpus Profiles

Corpus	Types	Tokens	Average of text	Number of Files
EILAC	11.530	137.853	690	200
EMLAC	15.339	295.922	1.479	200
ERDLAC	24.166	723.682	3.619	200
ECLAC	8.699	99.373	497	200
IILAC	12.563	181.086	905	200
IMLAC	8.116	117.414	588	200
IRDLAC	18.822	468.436	2.342	200
ICLAC	5.202	52.302	261	200

The profile of eight corpora showed by table 1 contain numbers of words that reflect quantity of native and non-native articles in language and linguistics subject category. In comparison, *English Introduction in Linguistics Article Corpus* (EILAC) has less numbers of text than *Indonesian Introduction in Linguistics Article Corpus* (IILAC) but the other three English corpora in method (EMLAC), research and discussion (ERDLAC), and conclusion (ECLAC) contain more tokens than Indonesian corpora.

Analytical Procedures

The frequency-based approach implemented by computer software is used to identify lexical bundles as unit of analysis. The frequency of lexical bundles as linguistic feature show that their occurrence is not by chance, but there are patterns of use (Sinclair, 2004). Threshold is set before the lists of

bundles are extracted and further reduced based on exclusion criteria namely overlapping and context-dependent bundles. The goal of frequency analysis is the list of lexical bundles that can be compared across text or article sections. After the lists of bundles are gained, this study conducts the comparison across article sections and focuses on the analysis of frequency.

Threshold needs to be determined in the context of frequency, range, and numbers of bundles. 4-word bundles are the most selected length by numerous researchers because of its manageable size. In this study, 3 until 5-word bundles are the focus in order to acquire various and more numerous results. The other criterion is that the bundles must occur at least 10% in corpus with minimum 20 frequency (Chen & Baker, 2010; Hyland & Jiang, 2018). The lexical bundles generated by corpus software need to be refined to remove overlapping bundles and context-dependent bundles. The normalization of raw frequency extracted automatically from software is conducted for comparable purpose (Yuliawati, 2018).

This study uses AntConc 3.5.9 (Anthony, 2020) as tool to analyse large number of words in corpora. It is one of the corpus software mostly used by studies of lexical bundles to analyse corpora (Bychkovska & Lee, 2017; Hyland & Jiang, 2018; Kwary et al., 2017; Sadat & Moini, 2014; Shin & Kim, 2017; Wright, 2019). It generates automatically bundle lists with adjustable threshold to set the minimum of frequency and range in clusters or n-grams tool.

FINDINGS AND DISCUSSION

Findings

In this section, the relative frequency of lexical bundles have been calculated automatically and the range of every bundle is displayed to see the distribution of bundles across corpora. The top 20 bundles in list are selected to discuss because they can represent the most commonly used bundles with high frequency and range in a particular corpus. The most frequent bundles in each text section are displayed by tables based on the rank. The relative frequency reflects the occurrence of a lexical bundle in corpus. The bundle *the use of*

displayed by table 3 indicates that this bundle occurs 128 times in a hundred thousand words. The range shows the amount of texts that use the bundle. To find the typical lexical bundles in a particular corpus, Microsoft excel is employed to highlight the duplicate values in lists to mark the same bundles. The analysis of frequency is conducted simultaneously with comparison between English and Indonesian corpus displayed by tables.

Table 3. List of lexical bundles in corpus of introduction

Indonesian Introduction (IILAC)				**English Introduction (EILAC)**		
Rank	Rel. Freq	Range	Lexical Bundles	Rel. Freq	Range	Lexical Bundles
1	128,116	85	the use of	64,562	63	as well as
2	54,118	69	as well as	54,406	53	the use of
3	49,700	55	in terms of	47,877	48	in order to
4	43,626	60	based on the	41,348	36	in terms of
5	41,969	47	in order to	38,447	40	one of the
6	32,581	49	is one of the	36,271	36	the development of
7	30,372	40	the process of	34,820	34	a number of
8	28,163	34	due to the	34,094	35	the role of
9	25,402	29	the implementation of	32,643	31	the field of
10	24,850	34	in other words	32,643	32	the present study
11	24,298	34	the development of	26,840	28	in the field
12	23,746	33	it can be	24,664	28	in this article
13	22,641	34	there is a	22,488	31	first language l
14	22,089	33	on the other hand	21,037	23	the current study
15	21,537	24	the results of	20,311	23	in relation to
16	20,985	29	in this study	19,586	23	in this study
17	20,432	28	the result of	19,586	25	the effects of
18	19,880	27	a number of	19,586	24	understanding of the
19	19,880	29	of the study	18,861	21	such as the
20	19,880	22	the ability to	18,861	20	the context of

Table 3 shows the identical patterns of use reflected by both corpora. The lexical bundles *the use of, as well as, in terms of, in order to* have the high scores in the bundle use in frequency and range. This authentic linguistic evidence become the marker of similarity between Indonesian and English writing in expert level. Apart from the similarity, there are the typical lexical bundles from different group of writers based on the computer calculation. In the corpus of Indonesian writing (IICLAC), the typical bundles are *the implementation of, the process of,* and *the ability to* that refer to the issue

concerned in the research. In EICLAC, the bundles *the field of, understanding of the,* and *such as the* are the typical bundles that cannot be seen in IICLAC. In the context of distribution, the bundles *the use of, as well as, in terms of* and *based on the* become the commonly used bundles in text introduction by both English and Indonesian writers. They reflect well-distributed bundles and used by more than fifteen writers in those corpora.

Table 4. List of lexical bundles in corpus of Method

				Indonesian method (IMLAC)		
English method (EMLAC)						
Rank	Rel. Freq	Range	Lexical Bundles	Rel. Freq	Range	Lexical Bundles
1	49,337	79	in order to	135,418	90	in this study
2	44,944	64	in this study	109,016	82	based on the
3	39,200	53	the number of	70,690	56	of this study
4	30,413	48	each of the	69,838	45	in order to
5	29,738	59	in terms of	63,877	51	of the study
6	29,062	56	one of the	58,766	52	the data were
7	29,062	56	the use of	57,915	41	the use of
8	27,034	54	a total of	48,546	36	the participants were
9	27,034	55	based on the	46,843	33	in terms of
10	27,034	46	the participants were	45,991	42	was used to
11	27,034	47	were asked to	40,029	34	in this research
12	24,331	49	included in the	40,029	39	this study was
13	21,965	50	of the study	38,326	29	one of the
14	21,965	29	of the target	38,326	28	the results of
15	20,951	41	in the study	38,326	34	this study were
16	19,938	44	of the participants	36,623	34	this study is
17	19,938	39	part of the	35,771	28	of the data
18	19,600	33	the present study	34,919	31	as well as
19	19,262	45	the end of	32,364	27	of this research
20	17,572	35	used in the	31,512	27	data from the

The bundles in the two lists showed by table 4 also provide the evidence that there is the presence of identical patterns of bundle use. Typical bundles in EMLAC are *included in the, the end of the,* and *a total of* that can be identical word combination in English method articles. IMLAC contains bundles *of the data, data from the,* and *in this research* with the relatively high range. Distribution of bundles in those two corpora show that the bundles *in this study*

and *in order to* are the most frequent multi-word unit that are used recurrently by English and Indonesian writers.

Table 5. List of lexical bundles in corpus of result and discussion

English Result and Discussion (ERDLAC)				**Indonesian Result and Discussion (IRDLAC)**		
Rank	Rel. Freq	Range	Lexical Bundles	Rel. Freq	Range	Lexical Bundles
1	30,262	89	in order to	76,425	128	based on the
2	27,084	70	the number of	31,808	69	in this study
3	26,393	76	in this study	31,381	67	most of the
4	21,142	45	the present study	30,100	70	in order to
5	20,589	83	one of the	28,606	74	related to the
6	19,069	57	in relation to	26,898	56	in the following
7	17,549	74	part of the	24,550	67	on the other hand
8	16,720	66	the role of	23,055	52	of this study
9	16,720	79	there is a	22,628	55	the form of
10	16,582	68	a number of	22,628	63	there is a
11	15,200	64	the importance of	22,202	56	the results of the
12	14,647	47	i don t	21,988	62	shows that the
13	14,509	75	based on the	21,775	49	as shown in
14	14,371	58	some of the	20,921	57	due to the
15	14,095	52	there was a	20,707	59	in other words
16	13,404	67	due to the	20,494	53	the findings of
17	13,127	49	the relationship between	20,280	52	the fact that
18	12,851	59	in addition to	20,067	27	of the word
19	12,298	41	the effects of	19,640	48	in the form of
20	12,160	56	can be seen	18,572	53	there is no

The table 5 above displays the corpora that contain the most numerous and various lexical bundles. There are numerous same bundles in the comparison because of the various patterns of bundle use. ERDLAC reflects typical bundles namely *there was a, can be seen,* and *the relationship between* that are not relatively frequent in the list. IRDLAC contains bundles *related to the, the fact that,* and *in the form of* in the top rank. In terms of distribution, the bundles *in this study, in order to, based on the*, and *one of the* become the familiar preference in both two group of writers.

Table 6. List of lexical bundles in corpus of conclusion

English Conclusion (ECLAC)				**Indonesian Conclusion (ICLAC)**		
Rank	Rel. Freq	Range	Lexical Bundles	Rel.	Rang	Lexical Bundles

				Freq	e	
1	6,944	52	as well as	15,869	47	the use of
2	5,333	37	in this study	10,133	41	of this study
3	5,032	38	in terms of	9,560	40	based on the
4	4,428	24	the current study	7,457	30	as well as
5	4,428	31	the use of	6,883	26	the results of
6	4,126	30	in order to	6,692	24	in terms of
7	3,824	29	of this study	5,927	20	the present study
8	3,522	30	the present study	5,162	24	in this study
9	3,522	25	the role of	4,589	21	of the study
10	3,220	22	in this article	4,398	21	it can be concluded that
11	2,818	25	some of the	4,015	21	due to the
12	2,616	20	a number of			
13	2,616	20	need to be			
14	2,616	21	one of the			
15	2,616	20	the development of			
16	2,415	23	the importance of			
17	2,214	20	for future research			

In these corpora, the lexical bundles displayed by table 6 are the least than the other three corpora (introduction, method, and result & discussion). It can be reasonable for the text length is the shortest. The bundles *as well as, in terms of,* and *the use of* are present in both corpora. The typical bundles *the current study, in order to,* and *the role of* become the most frequent in ECLAC that are not found in ICLAC. There is unpredictable result in ICLAC that the bundles *it can be concluded that* become the longest bundle in the top ten rank. This bundle can be the typical characteristic of Indonesian writers because it is familiar based on the statistical test. In the context of distributional analysis, bundles *as well as, the use of,* and *of this study* are well-distributed in both corpora.

Discussion

Based on the findings, the most numerous bundles occur across text sections are in the form of 3-word bundles which contain the most incomplete structure in this study. There are only 5 lexical bundles in 4-word length (*on the other hand, the results of the, in the form of, can be seen in, in the field of*) and one for 5-word length (*it can be concluded that*) in the top 20 rank. The

incomplete structure and the phrasal form of bundle investigated in this study can be the linguistic evidences that emphasize the use of phrase-based bundles.

The comparison between corpora reflects that the bundles across text section are identical. Although there are the same bundles used in both corpora, the typical bundles with high score of frequency and range are found to characterize the different group of writers. The typical lexical bundles found are not by chance but they indicate that there are patterns of bundle use in a group of writers and a particular discipline namely linguistics. The preference of writers creates the systematic patterns that can be identified in the form of lexical bundles.

The distributional patterns show that there is the presence of popular bundles in English and Indonesian writers. The top rank lists emphasize that the common lexical bundle structures are phrase-based in expert level. Both English and Indonesian expert level writers employ the phrasal bundles in their research articles. The list of the most commonly used bundles can be guidance of novice writers who want to improve their writing skill to acquire more acceptable writing style in research article.

CONCLUSION

The most frequent lexical bundles investigated by frequency criteria reflect the common pattern of bundle use in each corpus. Frequency-based approach to multi-word combination enables us to acquire reliable results because of its statistical test in authentic language data. The list of lexical bundles can be used for teaching and learning activities as well as the personal evaluation. Practically, this study can play role in English for Academic Purposes (EAP) to recommend prevalent patterns of lexical bundle use in the form of pedagogically useful list of word combination. The findings can also be used for non-native writers or scholars especially Indonesian writers to enrich the use of lexical bundles across sections in language and linguistics field.

REFERENCES

Anthony, L. (2020). *AntConc (Version 3.5.9).* Waseda University. https://www.laurenceanthony.net/software

Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and

written registers. *English for Specific Purposes*, *26*, 263–286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., & Reppen, R. (2015). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge University Press.

Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics*, *2*(Stubbs 2007), 229–253. https://doi.org/10.1075/ijcl.18.2.03bre

Budiwiyanto, A., & Suhardijanto, T. (2020a). Frequency and structure of Indonesian lexical bundles on academic prose in legal studies : A driven-corpus approach. *BASA*, *1999*, 1–7. https://doi.org/10.4108/eai.20-9-2019.2296703

Budiwiyanto, A., & Suhardijanto, T. (2020b). Indonesian lexical bundles in research articles : Frequency , structure , and function. *Indonesian Journal of Applied Linguistics*, *10*(2), 292–303. https://doi.org/10.17509/ijal.v10i2.28592

Bychkovska, T., & Lee, J. J. (2017). At the same time : Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, *30*, 38–52. https://doi.org/10.1016/j.jeap.2017.10.008

Byrd, P. A. T., & Coxhead, A. (2010). On the other hand : Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*, 31–64.

Chen, Y., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, *14*(2), 30–49.

Cortes, V. (2013). The purpose of this study is to : Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33–43. https://doi.org/10.1016/j.jeap.2012.11.002

Durrant, P. (2015). Lexical Bundles and Disciplinary Variation in University Students ' Writing : Mapping the Territories. *Applied Linguistics*, 1–30. https://doi.org/10.1093/applin/amv011

Farihah, L. Z & Rachmawati, E. (2020). Digital Hangman Game To Improve Student's Vocabulary Mastery In Teaching Narrative Text. Journal of Applied Linguistics and Literacy. Vol 4 (1), 38-46.

Gavioli, L. (2005). *Exploring Corpora for ESP Learning*. John Benjamins Publishing Company.

Hyland, K. (2008). As can be seen : Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*, 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K., & Jiang, K. (2018). Academic lexical bundles : How are they changing ? *International Journal of Corpus Linguistics*, *23 (4)*, 383–407.

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical Bundles in Journal Articles across Academic Disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 132–140. https://doi.org/10.17509/ijal.v7i1.6866

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals : Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, *21*, 60–71. https://doi.org/10.1016/j.jeap.2015.11.003

Sadat, Z., & Moini, M. R. (2014). Structure of Lexical Bundles in Introduction Section of Medical Research Articles. *Procedia - Social and Behavioral Sciences*, *98*, 719–726. https://doi.org/10.1016/j.sbspro.2014.03.473

Salazar, D. (2014). *Lexical Bundles in Native and Non-native Scientific Writing Applying a corpus-based study to language teaching*. John Benjamins Publishing Company.

Shin, Y. K., & Kim, Y. (2017). Using lexical bundles to teach articles to L2 English learners of different pro fi ciencies. *System*, *69*, 79–91. https://doi.org/10.1016/j.system.2017.08.002

Sinclair, J. (2004). *Trust the Text: Language, Corpus, and Discourse*. Routledge.

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews : Sections , frequencies , and functions. *English for Specific Purposes*, *54*, 1–14. https://doi.org/10.1016/j.esp.2018.09.001

Yuliawati, S. (2018). Perempuan atau Wanita? Perbandingan Berbasis Korpus Tentang Leksikon Berbias Gender. *Paradigma Jurnal Kajian Budaya*, *8*(2006). https://doi.org/10.17510/paradigma.v8i1.227

Yuliawati, S., Ekawati, D., & Erika Mawarrani, R. (2020). *Penulisan Akademik: Perspektif Linguistik Korpus dan Analisis Wacana*. UNPAD Press.