INDONESIAN JOURNAL OF APPLIED LINGUISTICS
Vol. 10 No. 2, September 2020, pp. 292-303

Available online at:
https://ejournal.upi.edu/index.php/IJAL/article/view/28592

https://doi.org/10.17509/ijal.v10i2.28592

Indonesian lexical bundles in research articles: Frequency, structure, and function

Adi Budiwiyanto and Totok Suhardijanto
Department of Linguistics, Faculty of Humanities, Universitas Indonesia, Jalan Prof. Dr. Selo Soemardjan, Depok, 16424, West Java, Indonesia

ABSTRACT

Recent studies show that lexical bundles in English are pervasively found in academic discourse. In addition, the characteristics of lexical bundles found vary and differ across registers and genres. Nevertheless, it is still interesting to carry out in languages other than English. This study aims to discover the characteristics of Indonesian lexical bundles that cover frequency, structure, and function in research articles. This study adopted a mixed-method. Identification of the lexical bundle was carried out using WordSmith 7.0 on a corpus comprising 3,125,546 words, taken from 1126 texts, and consisting of six disciplines. With a frequency threshold of 40 per million words and a minimum distribution of 5 texts, 197 lexical bundles have been obtained, consisting of three- to six-word bundles with a total occurrence of 51,813 times. In terms of structure, the incomplete structure is dominating the bundles by 78.7%, with a total frequency of occurrence 38,749 times. This research finds that the pattern of lexical bundles can be classified into five types: noun-based, prepositional-based, verb-based, adjective-based, and clause-based bundles. Lexical bundles in research articles are generally clause-based (49.2%). This indicates that Indonesian lexical bundles vary in structure. The use of clause fragments and passive verbs are the main features in this genre. In terms of the discourse function, research-oriented bundles are the functions that are commonly used, while participant-oriented bundles are the least. Each discourse function has its own structural characteristics. It is also found that one lexical bundle can have two functional categories. These findings contribute to a better understanding of the characteristics of written academic discourse. From the pedagogical point of view, the findings can be used as learning material for both native and non-native speakers.

Keywords: Discourse function; frequency; lexical bundle; structure; written academic discourse

First Received:	**Revised:**	**Accepted:**
29 December 2019	9 July 2020	19 August 2020
Final Proof Received:		**Published:**
21 September 2020		30 September 2020

How to cite (in APA style):
Budiwiyanto, A. & Suhardijanto, T. (2020). 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

INTRODUCTION

A research article is a prominent media for conveying ideas and knowledge to scientists and researchers (Hyland, 2009). For ideas and knowledge to be effectively conveyed to readers, proficiency in using standard words, phrases, and a formulaic language is needed. In other words, the ability to write for academic purposes is not only at the level of the lexicon and syntax but also requires a good ability to arrange a formulaic language which is very fundamental in written academic discourse. The series of formulaic sequence consists of basic elements in academic discourse and has specific structures and functions in certain fields of

* Corresponding Author
Email: adibudiwiyanto@gmail.com

science, registers, and genres. According to Coxhead and Byrd (2007, pp. 134-135), these sequences are important for writers and speakers for at least three reasons: 1) the word sets are often repeated and become part of the structural material used by advanced writers, making the students' task easier because they work with ready-made sets of words rather than having to create each sentence word by word; 2) as a result of their frequent use, such sets become defining markers of fluent writing and are important for the development of writing that fits the expectations of readers in academia; and 3) these sets of words often lie at the boundary between grammar and vocabulary and are so often revealed in corpus studies but much harder to see through analysis of individual texts.

This research focuses on the use of formulaic sequence, namely the Indonesian lexical bundle, in journal research articles. A lexical bundle is a type of formulaic language that has recently been widely studied. Wray (2002) defines a formulaic language as a series of words, both continuous (without being interrupted by other elements) or discontinuous (interrupted by other elements), which is prefabricated (i.e., stored in memory and can be recalled as a single unit when used or spoken) and is not produced or analyzed as separate units. Meanwhile, the research article is a genre that has attracted considerable attention from researchers (Cortes, 2013; Hyland, 2008, 2012; Jalali & Moini, 2014; Jalilifar et al., 2017; Kwary et al., 2017; Shahriari, 2017). Hyland (2012) states that a published research article is the most discursively crafted and rhetorically machined genre. It is characterized by lexical bundles that function to present research by engaging with literature, providing warrants, establishing background, connecting ideas, directing readers around the text, and specifying limitations.

The term *lexical bundle* was first used by Biber et al. (1999) in Longman Grammar of Spoken and Written English. They compared its usage in conversation and academic prose registers. Biber et al. (1999, p. 990) define lexical bundles as recurring sequences of three or more words, regardless of their idiomaticity, and regardless of their structural status. Lexical bundles are simply sequences of word forms that commonly go together in natural discourse. The bundles are identified by a frequency-driven approach. It means that there are a frequency and range threshold in the text. The frequency threshold indicates that the lexical bundles do not occur accidentally, while the range threshold indicates that the lexical bundles are not an idiosyncratic use of the individual speaker or writer.

Lexical bundles have been categorized in terms of their structures (Biber & Barbieri, 2007; Conrad & Biber, 2004; Hyland, 2008) as well as their functions (Biber & Barbieri, 2007; Conrad & Biber, 2004; Cortes, 2004; Hyland, 2008). Relating to their structure, only 15 percent of lexical bundles in conversation can be regarded as complete phrases or clauses, while less than 5 percent of the lexical bundles in academic prose represent complete structural units (Conrad & Biber, 2004; Cortes, 2004). Moreover, almost all the bundles bridge two structural units and are mostly not idiomatic. Hyland (2008) states that lexical bundles in an academic prose are generally in patterns of *preposition + nominal phrase fragment* (e.g., *in terms of the*, *at the end of the*), *nominal phrase + fragment phrase-of* (e.g., *the base of the*, *the structure of the*), or anticipatory *it* fragment (e.g., *it is possible to*, *it should be noted that*). The structure represents approximately 70 percent of the four-word bundles in written academic discourse and is rarely found in a conversation. Hyland (2012) also conducted a study comparing the use of lexical bundles in three genres, namely research article, dissertation, and thesis, consisting of four disciplines: electrical engineering, business studies, applied linguistics, and biology. The results show that text-oriented bundles (60.3%) are used most frequently in research articles, while participant-oriented bundles (14.2%) are the least used, and text-oriented bundles (25.5%) are in between.

Hyland (2008) and Salazar (2014) functionally classified the lexical bundles. The functions in their taxonomy refer to the meanings and purposes of the language. The functions try to organize the discourse according to situations or contexts. The three core categories in this taxonomy are 1) research-oriented bundles that help writers to structure their activities and experiences of the real world; 2) text-oriented bundles which are concerned with the organization of the text and its meaning as a message or argument, and 3) participant-oriented bundles that focus on the writer or reader of the text. Research-oriented bundles perform an ideational function; expressions in this category are location (e.g., *at the beginning of*), procedure (e.g., *was carried out*), quantification (e.g., *a large number of*), description (e.g., *the appearance of*), grouping (e.g., *this type of*), and topic (e.g., *the currency board system*). Text-oriented bundles are word combinations used to express textual functions. Some of the functions performed by these expressions are transition (e.g., *on the other hand*), comparative (e.g., *as compared with*), inferential (e.g., *these results suggest that*), causative (e.g., *as a result of*), structuring (e.g., *as described previously*), framing (e.g., *in the case of*), and objective (e.g., *to show that*). The following category, participant-oriented bundles, performs interpersonal functions. The functions performed by these expressions are stance (e.g., *is likely to*) and engagement (e.g., *it should be noted that*).

Studies on lexical bundles mostly focus on English. Nevertheless, it is still interesting to carry

out in languages other than English. At least it can be seen from the research conducted by Butler (1998), Cortes (2007), Tracy-Ventura et al. (2007) for Spanish and Kim (2009) for Korean. From their research, it is revealed that the occurrence of lexical bundles in a language is influenced by the structure of language and registers. In the context of the Indonesian language, we find lexical bundles in the Indonesian Web corpus of SketchEngine (https://app.sketchengine.eu), a general corpus with 90,120,046 words. Some of the top rank of lexical bundles with high frequency are *yang ada di* 'which exist in' (F=15,162), *oleh karena itu* 'therefore' (F=14,101), *dalam hal ini* 'in this case' (F=9,284), *yang dilakukan oleh* 'which is conducted by' (F=7,633), *yang berasal dari* 'which derive from' (F=7,036), dan *merupakan salah satu* 'is one of ' (F=6,982). Although the bundles are taken from the general corpus, we suspect that these bundles might belong to certain registers or genres and that some of them are common lexical bundles within academic disciplines.

As for the research on Indonesian lexical bundles, so far to our knowledge, it has not much been conducted. Samodra and Pratiwi (2018) investigated and compared lexical bundles of Indonesian and English on undergraduate thesis abstracts. They found that the Indonesian lexical bundles were dominated by phrase *penelitian ini* 'this research' and English bundles were dominated by the phrase *this research* 'penelitian ini'. Based on the structure, the Indonesian and English bundles are very similar in terms of word use, for example, *in this study*. Regarding the factors that influence the use of a lexical bundle, it is influenced by the author's knowledge of the rules of academic writing, language proficiency, and the differences in grammar rules of the two languages. It seems that their research is constrained by limited data. Therefore, the conclusion generated cannot be widely generalized to Indonesian lexical bundles. In terms of structure, Indonesian and English are different. The bundle *in the form of* when translated into Indonesian is realized in a word *berupa*, which is not a lexical bundle. Another example is the bundle *metode penelitian yang digunakan* becomes *the method that is* ***used in this research*** (where *used in this research* is a lexical bundle).

In summary, the aforementioned studies have improved our understanding of the use of lexical bundles in particular registers and genres. Besides, the results of those studies show that lexical bundles vary in terms of usage, structure, and function. Therefore, the present study investigates the use of Indonesian lexical bundles in research articles and attempts to answer the following research questions:

1. What lexical bundles are used in research articles?
2. How do lexical bundles vary within the six academic disciplines?
3. How do lexical bundles distribute in academic articles based on their structures and functions?

METHOD

This current research adopted a mixed-method (Cheng, 2012). It began with an exhaustive search for all three- to six-word lexical bundles. Then, it continued with observation to obtain regular patterns. Once patterns were found, tentative hypotheses were formulated so that they could be explored further and might develop into general conclusions (Biber, 2009). However, the process of functional classification required a top-down approach since the researcher had to consult the concordance lines to establish the functional categories of the lexical bundles. To implement this approach, a corpus consisting of a large number of texts was needed. The following is a description of the corpus used in this study.

Corpus design

The corpus for this study consists of research articles from six disciplines, namely medical science, nursing science, chemistry, computers, philosophy, and legal studies. The six disciplines were randomly chosen, and each of them represents different domains of research and methodological traditions. Medical and nursing sciences belong to the health domain; chemistry and computer belong to the science and computer domain, and philosophy and legal studies belong to the social and humanities domain. The research articles were taken from various journals published by universities or research institutions that are nationally indexed. Each field of the academic discipline consists of approximately 500 thousand words. Thus, the whole corpus comprises approximately three million words.

To assure that the corpus is representative, the articles used in this study were selected by stratified random sampling. The sample texts are texts published from 2010 to 2018 and have heterogeneous topics, volumes, and publishers. In addition, if the article is written individually, the author's name may only appear once. This is to avoid idiosyncrasy.

The texts obtained for this corpus are in the .pdf format. The texts were copied and then pasted into the MsWord document. The next step was to exclude bibliography, tables, charts/pictures, footnotes, headers and footers, the author's identity, and formulas. The clean texts were then saved in the .txt format with Unicode 8 (UTF 8) encoding. The files were then labeled consisting of science discipline, journal publisher, and article serial number. The following table is the size of the research article corpus.

Table 1
Corpus Size

Disciplinary Field_Text	Token (Running Words) in Text	Types (Distinct Words)	Type/Token Ratio (TTR)	Standardized TTR	STTR std.dev.	STTR Basis	
Medical science	235	528,106	27,146	5.14%	5.46%	91.68	1,000
Nursing science	202	522,068	23,665	4.53%	4.79%	92	1,000
Chemistry	243	528,747	25,795	4.88%	5.33%	91.2	1,000
Computer science	209	518,202	20,727	4.00%	4.38%	92.86	1,000
Legal studies	133	515,141	21,228	4.12%	4.29%	92.42	1,000
Philosophy	104	513,282	26,077	5.08%	5.24%	91	1,000
Total	1126	3,125,546					

Identification of lexical bundles

The lexical bundles examined are three-, four-, five, and six-word bundles. Lexical bundles are basically extended collocations based on the frequency of occurrence and the spread (or the range) of usage in the text (Biber et al. 1999, p. 992). Therefore, this study applied two criteria to identify lexical bundles, namely frequency, and range. The frequency threshold serves to prove that the lexical bundles are not accidental, while the range is to show that the bundles are not idiosyncrasies of particular speakers. This study set a cut-off frequency of 40 per million words, with a range of 5 texts. It means that if the corpus consists of 3,125,546 words, the frequency threshold used is 125 times, and it must appear at least in 5 different texts. WordSmith 7.0 (Scott, 2019) is used to extract the bundles.

Data analysis

To answer the questions in this study, frequency analysis of the lexical bundles was first carried out in the journals. Next, the structure was examined. Biber et al. (1999) showed that lexical bundles have strong grammatical correlations and produce a classification that groups them into several basic structural types. Then, functional analysis was carried out to classify lexical bundles into discourse functions. The classification adopted discourse functions of (Hyland, 2008) and (Salazar, 2014). They divide the function of lexical bundles into 1) research-based bundles, 2) text-oriented bundles, and 3) participant-oriented bundles. Meanwhile, to determine the functional categories of lexical bundles, the concordance of WordSmith 7.0 (Scott, 2019) was employed. The results are discussed in the following subsection.

FINDINGS AND DISCUSSION

The frequency of lexical bundles

Based on the predetermined identification parameters, it was identified that the corpus of research articles — consisting of 3,125,546 words— comprises 197 lexical bundles. They are three-word bundles (175), four-word bundles (18), five-word bundles (3), and six-word bundles (1). From the number of bundles, it indicates that the longer the lexical bundle, the less the number of occurrences. In this context, the ratio of occurence between bundles is quite large. A three-word bundle is almost as ten times as the number of a four-word bundle; a four-word bundle is as six times as a five-word bundle, and a five-word bundle is as three times as a six-word bundle.

From three-word to six-word bundles, they have a strong connection. A five-word bundle contains words that make up a four-word bundle, and a four-word bundle contains words that make up a three-word bundle. For example, the three-word bundle *dalam penelitian ini* is the element that makes up the four-word bundle *dalam penelitian ini adalah* and is the element of the five-word bundles *digunakan dalam penelitian ini adalah* and *yang digunakan dalam penelitian ini*. Another example is the bundle *dapat dilihat pada* becomes the element of the four-word bundle *dapat dilihat pada tabel,* and the bundle *pada penelitian ini* becomes the element of the four-word bundle *pada penelitian ini adalah*. In other words, the longer bundle is an extension of the shorter bundle (see Table 2). Moreover, there are also lexical bundles that are composed of a combination of two similar bundles. For instance, the six-word bundle *yang digunakan dalam penelitian ini adalah* is a combination of the five-word bundle *digunakan dalam penelitian ini* and *digunakan dalam penelitian ini adalah*.

In terms of frequency, the three-word bundles have the highest frequency of occurrence compared to the four-, five-, and six-word bundles. From Table 2, it can be seen that there are four bundles that occur more than 1000 times, namely *pada penelitian ini* (F = 1626/R = 558), *dalam penelitian ini* (F = 1418/R = 572), *penelitian ini adalah* (F = 1202/R = 563), and *oleh karena itu* (F = 1069/R = 534). The four bundles belong to text-oriented bundles. This indicates that those bundles play important roles in organizing the text and its meaning as a message or argument.

Table 2
The Top Lexical Bundles in the Research Article Corpus

No.	Three-word Bundle	Freq.	Texts	No.	Four-word Bundle	Freq.	Texts
1	*pada penelitian ini*	1,629	559	1	*dapat dilihat pada gambar*	545	197
2	*dalam penelitian ini*	1,420	573	2	*dalam penelitian ini adalah*	459	298
3	*penelitian ini adalah*	1,206	565	3	*penelitian ini bertujuan untuk*	347	290
4	*oleh karena itu*	1,070	535	4	*yang digunakan dalam penelitian*	321	226
5	*dapat dilihat pada*	988	362	5	*penelitian yang dilakukan oleh*	314	165
6	*merupakan salah satu*	890	541	6	*dapat dilihat pada tabel*	314	192
7	*yang dilakukan oleh*	881	409	7	*digunakan dalam penelitian ini*	312	223
8	*hasil penelitian ini*	731	313	8	*hal ini menunjukkan bahwa*	241	163
9	*yang digunakan adalah*	608	395	9	*pada penelitian ini adalah*	212	157
10	*dalam hal ini*	600	310	10	*hal ini sesuai dengan*	202	130
11	*dilihat pada gambar*	586	208	11	*hasil penelitian menunjukkan bahwa*	201	157
12	*yang digunakan dalam*	583	357	12	*penelitian ini menunjukkan bahwa*	185	125
13	*ini menunjukkan bahwa*	553	316	13	*ini bertujuan untuk mengetahui*	167	144
14	*penelitian yang dilakukan*	507	246	14	*hasil penelitian ini menunjukkan*	160	110
15	*yang digunakan untuk*	502	311	15	*penelitian ini adalah untuk*	157	145
16	*ini bertujuan untuk*	448	361	16	*hal ini disebabkan karena*	155	107
17	*dapat disimpulkan bahwa*	443	287	17	*tujuan penelitian ini adalah*	140	124
18	*penelitian ini bertujuan*	416	344	18	*dari penelitian ini adalah*	127	114
19	*yang berasal dari*	414	260				
20	*yang terdiri dari*	404	264				
No.	**Five-word Bundle**	**Freq.**	**Texts**	**No.**	**Six-word Bundle**	**Freq.**	**Texts**
1	*yang digunakan dalam penelitian ini*	299	215	1	*yang digunakan dalam penelitian ini adalah*	197	150
2	*digunakan dalam penelitian ini adalah*	200	152				
3	*penelitian ini bertujuan untuk mengetahui*	144	126				

The corpus of research articles comprises six sub-corpus, namely medical science, nursing, chemistry, computers, legal studies, and philosophy. Each discipline has lexical bundles that characterize those fields of sciences. Of the six fields, there are 16 shared lexical bundles that appear in all six fields/subcorpus (see Table 3). The bundles consist of 15 three-word bundles and one four-word bundle. These bundles are core lexical bundles on research articles.

Table 3
Shared Lexical Bundles

No.	Lecical Bundle	No.	Lexical Bundle
1	*dalam hal ini*	9	*sebagai salah satu*
2	*dalam penelitian ini*	10	*yang ada di*
3	*dalam penelitian ini adalah*	11	*yang berasal dari*
4	*dapat disimpulkan bahwa*	12	*yang digunakan dalam*
5	*hal ini dapat*	13	*yang digunakan untuk*
6	*merupakan salah satu*	14	*yang dilakukan oleh*
7	*oleh karena itu*	15	*yang lebih tinggi*
8	*penelitian ini adalah*	16	*yang terdiri dari*

From the table above, it shows that Indonesian bundles are all not idiomatic in meaning. The meanings are transparent from the individual words. In addition to the non-idiomatic meaning, the bundles also have characteristics in a structure. Some bundles, such as *dalam hal ini*, *dalam penelitian ini*, dan *oleh karena itu*, have complete structure, i.e., prepositional phrases. On the other hand, lexical bundles, such as *dapat disimpulkan bahwa*, *merupakan salah satu*, *sebagai salah satu*, dan *yang digunakan dalam*, are bundles with incomplete structure, i.e., there are fragmented parts. A more detailed explanation of this structure will be discussed in the following subsection.

The structure of lexical bundles

One of the characteristics of lexical bundles lies in its structure. Some studies show that lexical bundles generally have incomplete structures in a written register. Similarly, the lexical bundles in this research article corpus generally have incomplete structures. The incomplete structure is in the form of clauses, both free and bound clauses, which are fragmented on certain elements, such as the loss of an object, complement, or subject-complement, at once. The following are some examples.

(1) ***Hal ini menunjukkan bahwa*** *kriteria utama dalam kriminalisasi ialah berkaitan*

dengan aspek nilai-nilai moral yang ada dalam masyarakat. (HH040826)

(2) ***Tujuan penelitian ini adalah*** *untuk mengetahui potensi antioksidan ekstrak biji duku.* (SKI041543)

(3) *Mendongeng* ***merupakan salah satu*** *aktivitas yang dapat digunakan dan sesuai dengan perkembangan umur mereka*. (KP041843)

In the examples above, there are four lexical bundles, namely *hal ini menunjukkan bahwa*, *merupakan salah satu*, *tujuan penelitian ini adalah*, and *bertujuan untuk mengetahui*. In (1) it appears that the bundle is fragmented at the object slot; in (2) the bundle is fragmented at the complement slot; meanwhile, in (3) the bundle is fragmented at the subject-complement slot. They also serve as a bridge for two units, namely, the last word of the bundle becomes the first element of the next unit. For instance, the word *bahwa* in *hal ini menunjukkan* ***bahwa*** is the beginning of a nominal clause ***bahwa*** *kriteria utama dalam kriminalisasi ialah berkaitan dengan aspek nilai-nilai moral yang ada dalam masyarakat*; and the phrase *salah satu* in *merupakan* ***salah satu*** is the beginning of the phrase ***salah satu*** *aktivitas*.

(4) *Regenerasi* ***yang berasal dari*** *hepatosit matur berlangsung jauh lebih cepat dibandingkan regenerasi oleh sel oval.* (Jurnal-UI-2015)

(5) *Peralatan* ***yang digunakan dalam*** *penelitian ini meliputi peralatan gelas kimia standar.* (Jurnal-UIN-2013)

The incompleteness of the lexical bundles can be found not only in the independent clauses but also in the dependent clauses as in (4) and (5) above. The bundle *yang berasal dari* is a relative/adjective clause that is fragmented from its complete form *yang berasal dari hepatosit matur* and so is the bundle *yang digunakan dalam* that is fragmented from the complete form *yang digunakan dalam penelitian ini*. Such forms are commonly found in this corpus.

In the phrase level, there are also incomplete forms, such as *pada panjang gelombang, dengan hasil penelitian,* dan *seperti pada gambar*, as in (6)-(8). The complete structure of these bundles is successively *pada panjang gelombang maksimum 495 nm*, *dengan hasil penelitian sebelumnya*, dan *seperti pada gambar berikut ini*.

(6) *Pengukuran absorbansi dilakukan setiap 15 menit* ***pada panjang gelombang*** *maksimum 495 nm.* (SKI041318)

(7) *Temuan ini sesuai* ***dengan hasil penelitian*** *sebelumnya yang dilakukan oleh Whitfield, dkk.* (*2005*). (KP041810)

(8) *Tampilan antarmuka untuk login* ***seperti pada gambar*** *berikut ini*. (SKO040811)

The incomplete structure, in the form of both phrase and clause, can be found in quite large numbers in this corpus, i.e., 78.7% with a total frequency of occurrence at 38,749 times. Meanwhile, the rests are complete structure (21.3%) with a frequency of 13,064 times and are generally in the form of prepositional phrases, such as *oleh karena itu*, *dalam hal ini*, *dengan kata lain*, *di sisi lain,* and *pada tabel 1*. These findings appear to be in line with what by Biber et al. (1999) found, i.e., in academic writing, the lexical bundles generally have incomplete structure, and only 5% have complete structure. If compared to this study, there is quite a big difference. In fact, the difference is not equal in terms of the size of words that construct the bundles. The range in this study is three to six words, while Biber et al. (1999) focused only on four-word bundles. If the same analysis of the four-word bundles is carried out, it turns out that, from the 18 existing bundles (see Table 2), the whole form is definitely an incomplete clause. In other words, the results are not much different from Biber's (2009). It seems that incomplete structures of the Indonesian lexical bundle in academic writing have the same tendency to those in English.

The grammatical pattern of lexical bundles

After knowing the structure, it will also be interesting to go further on the grammatical pattern of the bundles. Based on the core elements that dominantly incorporate the bundles, they can be patterned into five types: noun-based, prepositional-based, verb-based, adjective-based, and clause-based bundles. The lexical bundles in research articles are mostly clause-based (49.2%), while the others have almost a similar number: noun-based (15.7%), prepositional-based (17.3%), verb-based (14.2%), and adjective-based (3.6%). Overall the use of clause-like bundles is higher than that of the phrase(-like) bundles. This indicates that article journal writers prefer using clause-based bundles to add or limit the topics or information. The patterns in detail can be seen below.

1. *Noun-based bundles*

The core elements of the bundles are nouns or noun phrases. The noun phrases can be formed by extending the noun to its right and/or left. The extension elements can be in the form of a clause or clause fragment that serves as a modifier or complementary.

- **NP (fragments) + *yang*-Clause fragments**

(9) ***Data yang digunakan*** *adalah data sekunder dengan menggunakan bahan hukum primer, sekunder dan tersier.* (HF0406091)

(10) *Uji statistik Manova terhadap data menunjukkan terdapat* ***perbedaan yang***

signifikan *antarkelompok perlakuan dalam jumlah sel ginjal rata-rata.* (SKI041542)

2. ***Prepositional-based bundles***

The core elements of the bundles are prepositions and noun phrases. The noun phrases that follow the preposition can be in the form of an incomplete noun phrase or complete noun phrase.

- **Prep + NP (fragments)**

(11) *Hal tersebut sesuai* ***dengan hasil penelitian*** *Chudlori dan Verma.* (KD040141)

(12) ***Oleh sebab itu****, budaya yang dikembangkan adalah budaya kepatuhan dan ketakutan.* (HF0404071)

3. ***Verb-based bundles***

The core elements of the bundles are verbs. The verbs can be extended by adding other elements after or before the verbs.

- **VPpassive + PP fragments**

(13) *Hal ini* ***dapat dilihat dari*** *angka laporan KDRT yang masuk ke pengadilan negeri, ...* (HH040411)

(14) *Hak moral ini secara eksplisit* ***diatur dalam pasal*** *24 UUHC.* (HH040815)

4. ***Adjective-based bundles***

The core elements of the bundles are adjectives and adverbs. The adverbs serve as modifying elements to the adjectives and can be located after or before the adjectives.

- **AP + PP fragment**

(15) *Karena logam berat dapat terakumulasi sedimen, maka kadar logam berat pada sedimen* ***lebih besar dari*** *air.* (SKI041509)

(16) *Pengaturan outsourcing* ***tidak sesuai dengan*** *teori hukum.* (HH040810)

5. ***Clause-based bundles***

The core elements of the bundles are clauses. A clause is a construction that contains a predicate and a subject with or without object, complement, or adverbial. The clause is either independent or dependent. An independent clause is one that can occur alone as a sentence, while a dependent clause cannot occur alone but is always part of a larger structure. It may occur embedded within a lower-level structure, such as a noun phrase.

- **Yang + VPpassive + PP fragments**

(17) *Daya listrik* ***yang dihasilkan dari*** *0,10 gram AF adalah sebesar adalah sebesar 45 -51 watt dengan daya rata-rata 0,083 watt per detik.* (SKI041119)

(18) *Bahan hukum primer adalah bahan-bahan hukum yang mengikat, terdiri dari peraturan perundangan* ***yang terkait dengan*** *obyek penelitian.* (HH040741)

- **Yang + A(P) + PP fragments**

(19) *Proses stimulasi membutuhkan media dan sarana belajar, seperti tersedianya alat permainan* ***yang sesuai dengan*** *usianya.* (KP040125)

(20) *Untuk proses dekompresi, tahapan dimulai dengan membuat layar citra baru dengan resolusi* ***yang sama dengan*** *citra asli.* (SKO040101)

- **NP + VP + (bahwa-Clause fragments)**

(21) ***Hasil penelitian menunjukkan bahwa*** *ada beberapa kesamaan dan perbedaan antara teori kebenaran Mo Tzu dan Pancasila.*(HF0404053)

(22) ***Ini menunjukkan bahwa*** *manusia memiliki kesempurnaan dalam penciptaannya*. (HF0406081)

- **NP (fragments) + AP + PP fragments**

(23) *Hal* ***ini sesuai dengan*** *penelitian yang menunjukkan bahwa paparan radiasi ultraviolet dapat menekan sistem imun.* (KD040827)

(24) *Hasil* ***ini sejalan dengan*** *penelitian yang dilakukan oleh American Academy of Sleep Medicine ....* (KP040166)

- **NP (fragments) + Vpassive + (PP fragments)**

(25) *Penelitian* ***ini dilakukan dengan*** *cara membagi budaya Minang ke dalam dua kelompok budaya besar, yaitu Minang Pesisir dan Minang Bukit.* (HF0404045)

(26) ***Hal ini disebabkan*** *adanya kehadiran kontaminan yang menggangu pertumbuhan sel Chlorella,sp.* (SKI041106)

- **NP (fragments) + V + untuk-Clause fragments**

(27) *Penelitian* ***ini dilakukan untuk*** *melihat gambaran dan perubahan histologis tubulus ginjal dihubungkan dengan dosis yang digunakan.* (KD040163)

(28) *Penelitian* ***ini bertujuan untuk*** *memberikan kontribusi dalam dunia game, khususnya game yang bersifat edukatif.* (SKO041614)

- **V(P) + NP + PP fragments**

(29) ***Tidak ada perbedaan*** *signifikan (p=1) lama rawat inap antara suplementasi zink saja atau dengan kombinasi probiotik.* (KD040818)

(30) *Tidak* ***ada hubungan antara*** *tingkat stres dan karakeristik pelaku rawat informal (95% CI; p >0.05*). (KP040168)

From the examples above, it indicates that there are typical patterns in research articles. *First*, passive bundles dominate in the corpus. In this context, a passive bundle is a means to present action and event by assuming that the actions and the events are the objects. *Second*, bundles with relative pronouns *yang* occur in a large number. These bundles are relative clauses and are widely used to provide additional explanations or compact the information on the subject, object, or complementary elements. *Third*, the inversion pattern is used in clause-based bundles, as shown in (29) and (30). It is usually used to introduce topics, i.e., *perbedaan* (29) and *hubungan* (30).

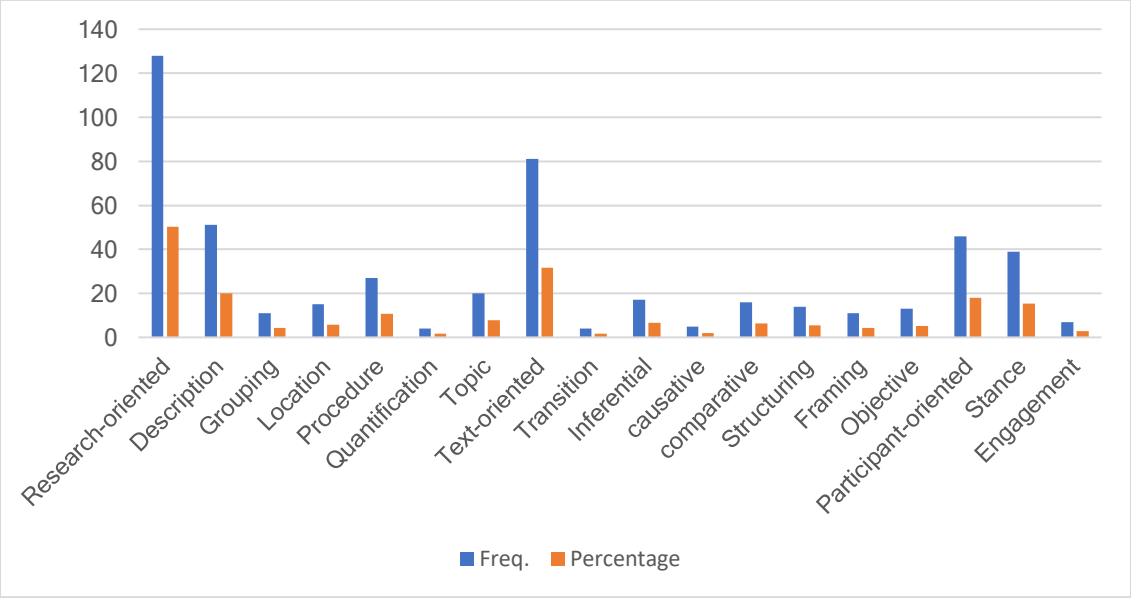
These patterns can be used, among other things, to organize the activities and experiences of the writer regarding his research, to organize the text and its meaning as a message or argument, and to focus the writer or the reader on the text. For more detail, the discussion on the function of lexical bundles will be given in the following subsection.

The functions of lexical bundles

This study found that research-oriented bundles are the most frequent function (50.2%), then followed by text-oriented bundles (31.8%). Participant-oriented bundles are the least bundles (18%). This suggests that research articles put more emphasis on presenting situations and events in the research as well as the entities, actions, and processes involved. In research-oriented bundles, the description bundles that are used to describe quality, condition, and existence are the highest function of use (20%). Meanwhile, in text-oriented bundles, the most frequently used function is the inferential marker (6.7%). The last, for the participant-oriented bundles, the stance function (15.4%) is more dominant than the stance function (15.3%). For more detail, the distribution of functional categories can be seen in Figure 1.

Figure 1

Distribution of Functional Categories and Subcategories



Research-oriented bundles

As mentioned earlier, this category is dominated by bundles that provide descriptions or explanations, whether they are objects, models, equipment, or research materials. Bundles with this description function are generally expressed in clause-based bundles, especially in the pattern of *yang + VP / AP + PP fragment*, as seen in the following examples.

(31) *Konsep dasar terjadinya DA adalah melalui reaksi imonologi yang diperantarai oleh sel-sel* ***yang berasal dari*** *sumsum tulang.* (KD041708)

(32) *Akan tetapi, tidak semua objek* ***yang ada di*** *alam merupakan fraktal.* (SKO040113)

(33) *Dari perspektif hi[e]rarki Peraturan Perundang-undangan, peraturan* ***yang lebih rendah*** *tidak boleh mengubah peraturan yang lebih tinggi.* (HH040807)

In addition to the description function, bundles with the procedure function are also commonly found, i.e., 11.8%. This function shows events, activities, and methods of the research. The procedure bundles generally use verb-based bundles, especially the passive structure, to show the research process or the activities, as shown below.

(34) *Larutan tersebut kemudian* ***dimasukkan ke dalam*** *gelas piala dan diuapkan dalam waterbath pada suhu 50°C sampai menjadi kering.* (SKI041541)

(35) *Pengurangan fosfat banyak* ***dilakukan dengan cara*** *adsorpsi yakni menggunakan selulosa dalam biomassa.* (SKI041541)

As for the other research-oriented bundles, namely location, quantification, grouping, and topic, they appear in a small number. Even though they are small in number, the bundles still contribute to the accuracy of the research process by identifying location and orientation (36), determining size and number (37), showing groups or parts (38), and showing the subject of research (39). The location function is often realized in the form of prepositional-based bundles. The grouping function is usually expressed in verb-based bundles. Meanwhile, the quantification and topic functions are generally manifested by noun-based bundles.

(36) ***Di rumah sakit*** *bakteri biasanya ditularkan melalui petugas, alat-alat atau pengobatan parenteral.* (KD040141)

(37) ***Sebagian besar responden*** *memiliki pengetahuan yang buruk terkait suhu minimal untuk pemberian obat atau melalui mandi seka*. (KP041547)

(38) *Banyumas* ***merupakan salah satu*** *wilayah yang menyumbang angka prevalensi penderita talasemia yang cukup besar*. (KD040842)

(39) *Praktik pembentukan hukum seperti ini tidak sesuai dengan prinsip-prinsip* ***hak asasi manusia***. (HH040724)

The extensive use of research-oriented bundles in research articles indicates that this kind of genre places more emphasis on research practice and the methods, procedures, and equipment used as well as the research objects.

Text-oriented bundles

The function of this bundle relates to the organization of texts and their meaning as a message or argument (Hyland, 2008). There are two subfunctions that dominantly appear in this bundle, namely inferential (6.7%) and comparative (6.3%). The inferential functions are associated with drawing conclusions from data. They are usually expressed by clause-based and verb-based bundles, as in the following examples.

(40) ***Hasil penelitian ini menunjukkan bahwa*** *pagelaran wayang kulit purwa adalah suatu karya seni fenomenal yang memiliki kisah-kisah yang dapat dijadikan gambaran bagi kehidupan manusia.* (HF0404049)

(41) *Dengan demikian* ***dapat disimpulkan bahwa*** *tindakan pencegahan penularan HIV oleh ODHA di Sorong dapat diprediksi oleh variabel umur.* (P040170)

(42) *Dari uraian di atas* ***dapat dilihat bahwa*** *argumen Indonesia untuk menyatakan bahwa virus merupakan sumber daya hayati yang tunduk pada ketentuan CBD tidak dapat dipertahankan.* (HH040416)

The comparative function deals with comparing and contrasting different elements. This function is often realized in clause-based bundles; that is (NP) + AP + PP fragment, as shown in the following example.

(43) *..., tetapi aspirasi* ***agama lebih tinggi*** *daripada filsafat.* (HF0404062)

(44) ***Hal ini sesuai dengan*** *teori Health Belief Model yang menunjukkan bahwa persepsi seseorang akan memengaruhi tindakan seseorang.* (KP041830)

(45) *Hal* ***ini sejalan dengan*** *asas legalitas dan asas lex certa.* (HH040909)

In addition to the two functions above, structuring functions (46), objective (47), and framing (48) also appear quite frequently in research articles. The structuring function relates to reflexive text markers that organize a text, sequences or direct the reader to a specific place in the text. This function usually uses prepositional-based bundles. The objective function relates to the author's purpose and is usually indicated by clause-based bundles. Meanwhile, the framing function is associated with the conditioning of arguments by specifying condition boundaries. This function is usually represented by prepositional-based bundles.

(46) *Secara garis besar, permainan ini memiliki alur kerja* ***seperti pada Gambar*** *8.* (SKO040707)

(47) ***Penelitian ini bertujuan untuk*** *mengetahui kebijakan pengupahan terkait dengan investasi di Indonesia.* (HH040928)

(48) ***Dalam hal ini*** *ilmu tidak dapat dilepaskan dengan keberadaan metodologi dalam pencapaian pengetahuan.* (HF0404050)

Other functions that appear with very low numbers are the transition function (49) and causative (50). Although the level of occurrence is the lowest, their role in uniting ideas in the text is very important. Transition bundles are usually characterized by prepositional phrase structures, whereas causative bundles are usually in the form of clause fragment structures.

(49) ***Hal ini disebabkan*** *oleh kecepatan difusi partikel kolesterol melalui pori-pori karboksimetil kitosan yang cepat.* (SKI041541)

(50) ***Oleh karena itu,*** *para pemikir feminis berusaha memberikan sanggahan untuk menolak konsep ketubuhan patriarkis.* (HF0404074)

From the description above, it is obvious that the bundle with this function is very helpful for the writer in producing a unified and integrated idea. With this bundle, the writers are able to communicate the interpretation of their data, and the readers at once facilitate the process of reading articles through arguments that are arranged in a structured and logical manner. All these functions form the foundation of effective argumentation.

Participant-oriented bundles

This functional category deals with two-way interactions between the participants in the text, namely the writer and the reader. By expressing the epistemic, evaluative, and directive meaning, participant-oriented bundles help writers convey their attitudes towards their assertions and establish the appropriate relationship with their readers (Hyland, 2005). There are two functions in this category: stance and engagement. The stance functions are to convey the writer's attitudes and evaluations. They can generally be expressed by verb-based bundles or clause-based bundles, as can be seen below.

(51) *Temuan penelitian ini **dapat digunakan sebagai** acuan oleh praktisi keperawatan untuk mengembangkan cara penanganan ketidakpatuhan klien skizofrenia.* (KP040122)

(52) *Analisis transpor **dapat dilakukan dengan** beberapa metode, misalnya metode lag time difusi dan metode pemodelan kompartemen.* (SKI041137)

(53) ***Hal ini menunjukkan bahwa** kemampuan interaksi antara zeolit termodifikasi ditizon lebih cepat dibandingkan dengan zeolit saja.* (SKI041010)

The engagement function relates to the way the writer recognizes the presence of the reader rhetorically to invite/actively attract the reader along with the arguments that the authors convey, include them as involved in discourse, and guide them in interpreting (Hyland, 2005). The engagement functions are generally directive, and which are expressed through verb-based bundles. They direct the reader to engage in textual and cognitive activities. This function can be seen from the following bundles.

(54) *Distribusi frekuensi umur, jenis kelamin, tingkat pendidikan dan lama kerja **dapat dilihat pada** Tabel 1*. (KD040408)

(55) *Hasil peremajaan, inokulasi dan produksi **ditunjukkan pada Gambar** 1*. (SKI041207)

(56) *Hal **ini dapat dilihat** dalam Pasal 1*. (HH040402)

(57) *Peningkatan kemampuan ibu **dapat dilihat dari** peningkatan pengetahuan, sikap dan perilaku dalam menstimulus balita.* (KP040125)

Examples (54) to (57) comprise lexical bundles with engagement functions in the form of textual activities that direct the reader to another part of the text or other texts. Meanwhile, (61) shows cognitive activity that directs the readers to interpret an argument or encourage them to note, acknowledge, or consider an argument.

Multifunctionality of lexical bundles in research articles

In several studies of lexical bundles (Hyland, 2008; Jalilifar, Ghoreishi, & Roodband, 2017; Salazar, 2014) it is found that a lexical bundle may perform more than one function in different contexts. Similarly, this study found such multifunctionality. There are 58 lexical bundles that have multiple functions. For example, the bundle *dalam penelitian ini*, in addition to functioning to indicate a place, as shown in (58), it can also serve as text reflexive markers that organize discourse, as shown in examples (59) and (60). Changes of the functional category are affected by the position of the bundles in a sentence.

(58) *Inilah keterkaitan antara kajian ontologi dengan kajian politik sebagaimana terlihat **dalam penelitian ini**.* (HF0404054)

(59) *Besar nilai fitness **dalam penelitian ini** sangat dipengaruhi oleh besarnya ukuran populasi.* (SKO040833)

(60) ***Dalam penelitian ini** digunakan 500 data citra yang dibagi menjadi 5 kategori, ...* (SKO040841)

In addition, inferential bundles and stance bundles, as well as comparative bundles and position bundles, may also perform a relationship that is similar to location bundles and framing bundles, as shown in the following examples.

(61) ***Hal ini menunjukkan bahwa** ada perbedaan signifikan kadar MDA dan GSH antara kelompok I, kelompok II, kelompok III dan kelompok IV*. (SKI041543)

(62) ***Hal ini sesuai dengan** rumusan yang terdapat dalam Pasal 33 ayat (1) dan ayat (4) UUD 1945*. (HH040816)

CONCLUSION

The results suggest that the Indonesian lexical bundles in research articles have their own characteristics. They include frequency, structure, and discourse function. In terms of frequency, there are 197 lexical bundles consisting of three to six words with a total occurrence of 51,813 times. A three-word bundle is the most common bundle, while a six-word bundle is the least one. From the corpus consisting of six academic disciplines, it is

found that there are 19 core lexical bundles, i.e., bundles that appear in all six disciplines.

In terms of structure, the lexical bundles can be categorized into complete and incomplete structures. The incomplete structure can be found in the form of clauses and phrases. This incomplete structure is dominating the bundles by 78.7%, with a total frequency of occurrence 38,749 times. In addition to incomplete structures, complete structures are also found, and they are generally in the form of phrases. The pattern of lexical bundles can be classified into five types: noun-based, prepositional-based, verb-based, adjective-based, and clause-based bundles. Lexical bundles in research articles are generally clause-based (49.2%). The use of clause fragments and passive verbs is the main feature in this genre.

In terms of the discourse function, research-oriented bundles are the function that commonly appears. This function relates to how the authors compile their activities and experiences regarding their research. Meanwhile, the least used are participant-oriented bundles, which focus on the writer or reader of the text. Each discourse function has its own structural characteristics. In other words, grammatical patterns can show a particular function of a lexical bundle. The analysis also found that one lexical bundle can have two functional categories.

The findings in this study contribute to a better understanding of the characteristics of written academic discourse. From the pedagogical point of view, the findings can be used as learning material, for both native and non-native speakers. For many Indonesian language learners, one of the difficulties faced is collocation. By studying lexical bundles, it means that they are also studying collocations because lexical bundles are extended collocations.

ACKNOWLEDGMENTS

This study was conducted with the financial support provided by the Directorate of Research and Community Engagement, Universitas Indonesia.

REFERENCES

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics, 14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Pearson.

Butler, C. S. (1998). Collocational frameworks in Spanish. *International Journal of Corpus Linguistics, 3*, 1–32.

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge.

Conrad, S. M., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica, 20*, 56-71.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, V. (2007). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 3*(1), 43–57.

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

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147. https://doi.org/10.1016/j.jslw.2007.07.002

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies, 7*(2), 173-192. https://doi.org/10.1177/1461445605050365

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21. https://doi.org/https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2009). *Academic discourse: English in a global context*. Continuum.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*, 150-169. https://doi.org/10.1017/s0267190512000037

Jalali, Z. S., & Moini, M. R. (2014). Structure of lexical bundles in the introduction section of medical research articles. *Procedia - Social and Behavioral Sciences, 98*, 719-726. https://doi.org/10.1016/j.sbspro.2014.03.473

Jalilifar, A., Ghoreishi, S. M., & Emam Roodband, S. A. (2017). Developing an inventory of core lexical bundles in English research articles: A cross-disciplinary corpus-based study. *Journal of World Languages, 3*(3), 184-203. https://doi.org/10.1080/21698252.2017.1301279

Kim, Y. (2009). Korean lexical bundles in conversation and academic texts. *Corpora, 4*(2), 135-165. https://doi.org/10.3366/e1749503209000288

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics, 7*(1). https://doi.org/10.17509/ijal.v7i1.6866

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. John Benjamins.

Samodra, M. C., & Pratiwi, V. D. R. (2018). Lexical bundles in Indonesian and English undergraduate thesis abstracts. *The Fourth Prasasti International Seminar on Linguistics Advances in Social Science, Education and Humanities Research, 166,* 193-191. https://doi.org/10.2991/prasasti-18.2018.35

Scott, M. (2019). *WordSmith tools* (Version 7.0). Stroud: Lexical analysis software. https://www.lexically.net

Shahriari, H. (2017). Comparing lexical bundles across the introduction, method, and results sections of the research article. *Corpora, 12*(1), 1-22. https://doi.org/10.3366/cor.2017.0107

Tracy-Ventura, N., Cortes, V., & Biber, D. (2007). Lexical bundles in Spanish speech and writing. In G. Parodi (Ed.), *Working with Spanish Corpora* (pp. 354–375). Continuum.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.