

Lexical Bundles in Indonesian and English Undergraduate Thesis Abstracts

Maria Caroline Samodra

Language Institute of Sanata Dharma University
Yogyakarta, Indonesia
s.mariacaroline@gmail.com

Vincensia Dian Ratna Pratiwi

English Language Studies, Sanata Dharma University
Yogyakarta, Indonesia
vincensia.dian@gmail.com

Abstract—The present study investigates Indonesian and English lexical bundles (LBs) used by Indonesian learners in their thesis abstracts. It aims to examine (1) the most common lexical bundles in the Indonesian and English undergraduate thesis abstracts, (2) similarities/differences in terms of form and functions, and (3) possible factors for the use of LBs in the students' abstracts. The data were obtained from a corpus of 140 abstracts written by Indonesian undergraduate students from Natural Science and Social Science disciplines in several universities in Indonesia. Results show that LBs with the phrases *penelitian ini* and *this research* dominate the abstracts. Based on the form, Indonesian and English LBs are mostly similar in the way that the same words are used repetitively from one language to the other, such as *dalam penelitian ini*, which often becomes in this research. Functionally, research-oriented and text-oriented LBs dominate the abstracts in both languages while the participant-oriented ones are the least frequent LBs. It is implied that the LB usage is influenced by students' knowledge of academic conventions, students' proficiency on the English LBs, and the differences of the languages themselves. There are some implications for teachers and students regarding the use of LBs in academic writing.

Keywords—lexical bundles, abstracts, undergraduate, corpus

I. INTRODUCTION

According to Swales (1990:179), an abstract functions as the “front matter” and “summary” of the research, making them a standalone genre. The importance of the abstract is highly significant in the way that many readers choose to read the abstract before deciding whether to read the content of the research. It is also often the case that abstracts are translated from one language to English to enable wider communities. Therefore, it can be implied that high-quality abstracts interest the audience through the language used.

An aspect of abstracts which might affect how they are structured is lexical bundles, or co-occurring words. Research has shown that the lexical phrases identified in academic writing have functions dependent on its sections (e.g. Ngadiman 2013 and Cortes 2013), suggesting that LBs are also essential for writing an abstract. However, such specific distinctions in academic writing have tended to be overlooked; the majority of the LB studies have only focused on academic writing in general (e.g. Hyland 2008; Jalilifar et al. 2017) although most have shown striking differences among different communities. Due to the importance of abstracts in

academic writing, this present study focuses on learners' use of LBs in their mother tongue (L1), Indonesian, and their second/foreign language (L2), English, in the undergraduate thesis abstracts to examine how abstracts are written in L2 in relation to their L1.

Three questions are addressed in the study, namely: 1) What are the most common lexical bundles in the Indonesian and English undergraduate thesis abstracts?, 2) How are the Indonesian and English lexical bundles similar/different in terms of form and functions? To what extent do the English LBs conform to the academic LB list?, and 3) What possibly influences the similarities/differences in the use of LBs in the Indonesian and English abstracts?.

II. REVIEW OF LITERATURE

Lexical bundles are important as they serve as building blocks for writing (Biber and Conrad 1999; Hyland 2008). Formulaic expressions have given rise to the increase of LB studies due to the advancement of corpus-based methodologies. Identifying the most frequent bundles enable the writers to analyse the discourse organisation (Cortes 2013: 35). In academic writing, knowing how to use fixed expressions can help promote cohesion of the texts.

Biber et al. (1999) define lexical bundles as three- or more word units that occur frequently and commonly go together. It is to be noted that the shorter bundles are often incorporated in the longer ones, for example: I don't think and well I don't think (ibid: 992). This definition has appeared to be the basis of other studies (e.g. Hyland 2008 & Cortes 2013).

Hyland (2008) explored 4-word bundles in a corpus of academic texts (e.g. research articles, PhD theses, and MA dissertations). Drawing from his data, he classified 3 functions of lexical bundles: (1) research-oriented, (2), text-oriented, and (3) participant oriented. Generally, the most frequent bundles were the first two functions as his own study, also supported by Wei and Lei (2011), revealed so.

Following Hyland (2008), Jalilifar et al. (2017) attempted to invent a general lexical bundle list from a 6-million-word corpus of research articles from prestigious international journals. Nevertheless, in contrast to Hyland (2008), they found that most of the lexical phrases were text-oriented.

Research show that lexical bundles are used differently between native (NS) and non-native (NNS) writers. Chen and

Baker (2010) investigated the LBs of native expert writing and non-native Chinese student writing. The results showed that the widest range of lexical bundles were available in native expert writing. Similarly, Adel and Erman (2012) studied NS and Swedish NNS writing and it revealed that the NS texts had more varied and wide-ranging bundles, for example existential there and unattended this.

Following the previous studies, Dontcheva-Navratilova (2012) analysed Czech students' diploma theses, which suggested that L1 transfer played an important role in the students' writing. The L1 interference on LBs was then confirmed by Paquot (2013), who discovered that learners' LB use could be traced back in their mother tongue, French.

Studies on lexical bundles in different languages also show cross-linguistic differences in how formulaic expressions are utilised. Granger (2014), for example, compared English and French parliamentary debates. The results indicated that the LBs of different languages varied in length, and that different genres have their own key bundles. For instance, the proportion of the 1st person pronoun in English parliamentary debate was significantly larger than that in French. Novita and Kwary (2018), in addition, investigated how students and professional writers translated LBs in a literary text from Indonesian into English. It was revealed that the proficiency level of the writer groups affected their production of LBs.

With the exception of Novita and Kwary (2018), there seems to be little attention to contrastive studies of lexical bundle among non-native speakers of English. Additionally, previous research on academic writing has tended to ignore the specific sections of academic papers, which, as Cortes (2013) discovered, influence the use of word combinations. Therefore, this present research aims to fill the gap by investigating the use of lexical bundles in Indonesian and English abstracts among Indonesian undergraduate students.

III. METHODOLOGY

This research used the corpus-based approach (McEnery et al. 2006: 10), which, in nature, is quantitative and qualitative. Quantitatively, the frequency counts of the English and Indonesian LB list were relied on to determine the most common bundles comprising each language's abstracts. Qualitative interpretation was conducted by analysing the LB list and comparing it to the pre-existing LB list by Jalilifar et al. (2017) and discussing LB functions identified in the abstracts.

A. Corpus

The data were taken from a total of 140 Indonesian and English abstracts in the undergraduate theses from various universities in Indonesia written in 2016-2018. The English abstracts were the translated version of the Indonesian abstracts to fit the aims of the present study. Analysing parallel texts, or the same texts written in different languages, as Baker (1995) suggests, might shed light to how writers compose in a foreign language in relation to their first language.

The discipline differences were not taken into account as the purpose of this research was to obtain core patterns of the bundles. However, for the data to be balanced, as McEnery et al. (2006: 19) suggest, respectively 36 and 34 texts from Natural and Social Sciences were chosen using a purposive random sampling technique. Table 1 shows the composition of the corpus of Indonesian and English abstracts.

TABLE 1. Corpus of Indonesian and English abstracts

Disciplines	Indonesian(TXT)		English (TXTRANS)	
	No. of text	No. of words	No. of text	No. of words
Natural Science	36	8175	36	8879
Social Science	34	7314	34	8241
TOTAL	70	15489	70	17120

B. Research Procedure

Adapted from Lee (2013), four steps were taken to answer the research questions.

1) 1. Extracting LBs from English and Indonesian abstracts

3-word and 4-word bundles were extracted from the corpus using AntConc, a concordancing software (Anthony 2014). The LBs which occurred more than 3 times in more than 3 abstracts were extracted to identify the representation of the most common bundles in the abstracts. The reason for choosing the cutoff point is to avoid idiosyncrasies of the writers (Biber et al. 2004).

2) Selecting LBs which are relevant to the research purposes

To obtain the general tendency of the word combinations, the discipline-specific bundles were eliminated, for example: *organic rice farming*. With such bundles eliminated, the list would reveal the core word combinations throughout the abstracts.

3) Organising the data

Similar to those of Hyland (2008) and Jalilifar et al. (2017), the LBs were first organised based on frequency ranks. Next, the functions were determined by categorising each bundle into research-, text-, and participant-oriented bundles (Hyland 2008). To ensure validity of the function categorisation, both researchers had to reach an agreement in coding the LBs.

4) Interpreting the data

Interpretation was made by observing the frequency ranks, similarities/differences, and functions. As the sizes of the two corpora are different, the frequencies were normalised to per 10,000 words to make them directly comparable. As suggested by McEnery and Hardie (2012), the formula for normalising the frequencies was:

$$F/10,000 = \frac{\text{LB raw frequency}}{\text{number of words in the subcorpus}} \times 10,000$$

Since Jalilifar et al.(2017) by far have created the most comprehensive English LB list in academic writing, the English LBs in the abstracts were compared to the list (see

Appendix 2) to see to what extent the students' L2 writing conforms to the English academic conventions.

IV. RESULTS AND DISCUSSION

A. Most Common Lexical Bundles

There are 551 Indonesian 3- and 4-word LBs (F/10,000=356.01) and 674 English LBs (F/10,000=393.69) identified in the abstracts. Due to space constraints, Table 2 (see Appendix 1) shows only the twenty most frequent bundles in Indonesian and English. Based on the table, it can be seen that the Indonesian and English LBs are highly similar in the way that the LBs are dominated by the phrases *penelitian ini* and *this research*. For example, the most common 3-word LBs in the Indonesian abstracts are *penelitian ini adalah* (F/10,000=29.72) and *dalam penelitian ini* (F/10,000=21.97). Similarly, the English counterpart this research is (F/10,000=21.03) and of this research is (F/10,000=10.51) are the most frequent 3- and 4-word bundles respectively.

B. Lexical Bundle: Forms and Functions

1) Forms of LB

The data show that there is a great number of Indonesian LB patterns retained in their English structures. For instance, the phrase *dalam penelitian ini* (1) possesses the same structure in the English version as in *this research* (2), noticed to be used repetitively in different texts.

- (1) Metode penelitian yang digunakan **dalam penelitian ini** adalah metode kualitatif. Pendekatan **dalam penelitian ini** adalah menggunakan pendekatan deskriptif kualitatif. Teknik pengumpulan data **dalam penelitian ini** yaitu observasi, wawancara dan dokumentasi. (TXT 21)
- (2) The approach **in this research** is qualitative descriptive approach. The collecting data techniques **in this research** are observation, interview and documentation. (TXTRANS 21)

However, compared to the LB lists by Jalilifar et al. (2017) and Hyland (2008), the phrase with *this research* is not a frequently used bundle by expert writers. Instead, it is more common to use *in this study* and *in this paper*, as indicated by its high rank in the list. Therefore, it is obvious that *this research* is idiosyncratic to Indonesian undergraduate writers of the current study.

Other frequently used LBs in the Indonesian abstracts are the ones with *untuk mengetahui*, as in *bertujuan untuk mengetahui* and *ini bertujuan untuk mengetahui*. Surprisingly, the English correspondences are somewhat similar through the use of *to know the*, *is to know*, *is to know the*, and *research is to know* as can be seen in (3), (4), (5), and (6). Although other variations such as *is to determine*, *is to describe* and *is to analyse* can also be identified, it was less frequent than the LBs with *to know*.

- (3) Penelitian **ini bertujuan untuk mengetahui** peran word of mouth dalam membentuk fanatisme masyarakat Kota Padang pada batu akik. (TXT 21)
- (4) The purpose of this research is **to know the** role [sic] of word of mouth in shaping Padang City society fanaticism to akik stone. (TXTRANS 21)
- (5) Tujuan penelitian **ini adalah mengetahui** gambaran happiness pada individu... (TXT 69)
- (6) The goal of this research is **to know the** picture of happiness in the three informants... (TXTRANS 69)

Those bundles are not common and natural in the English academic discourse since they do not exist in the academic LB lists (Jalilifar et al. 2017). Thus, it could be inferred that the phrase *untuk mengetahui* and the English counterpart *to know* are typical of Indonesian abstracts.

Apart from the similarities, a few LB differences are observed in the occurrence of lexical phrases with high frequencies in one language but low in the other. As an instance, *one of the*, which is also common in the academic LB list (Jalilifar et al. 2017), is the second most frequent LB in the English abstracts. Yet, the Indonesian LBs with the phrase *salah satu* (*merupakan salah satu*, *adalah salah satu*, and *sebagai salah satu*) have a relatively low rank, meaning that they are not commonly used in Indonesian. Example (7) shows how *one of the* is used in one of the texts.

- (7) **One of the idea** to distract children to watch animation movie (TXTRANS 15)

Besides (7), there are other incorrect forms in the English abstracts such as *one of effort* and *one of many technique*. In this case, the low rank of Indonesian LBs with *salah satu* might cause the wrong use of *one of* because they are not familiar with how it is used in English. In Indonesian, *salah satu* is not followed by plural forms; this might have contributed to their mistakes for not using plural forms following *one of*. In this case, language transfer, as Dontcheva-Navratilova (2012) and Paquot (2013) suggest, might contribute to ungrammatical bundles discovered in the data.

Another small proportion of LBs simply show differences between languages; they exist only in one language, but not in the other. For example, in (8) and (9), *in the form of* is realised as *berupa*, which is not a lexical bundle.

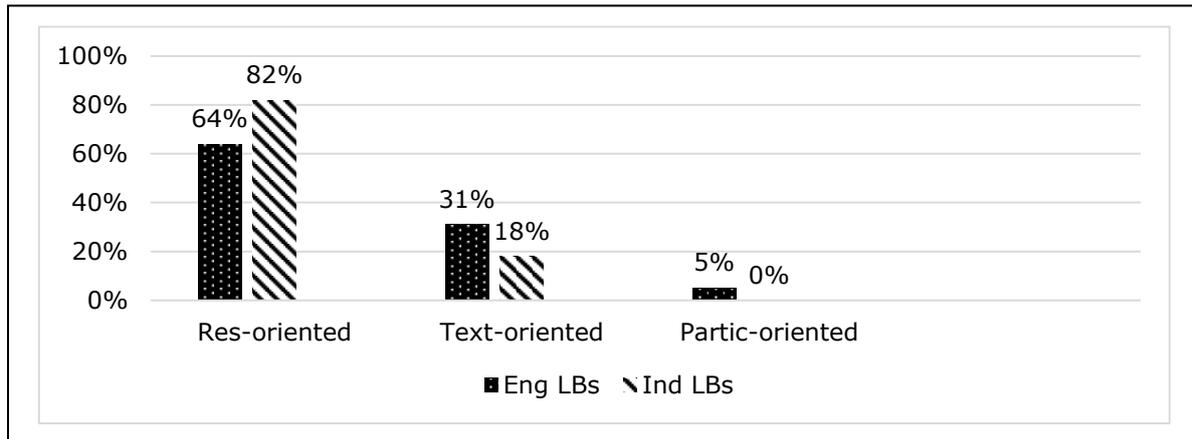
- (8) Hasil yang diperoleh **berupa** peta parameter-parameter yang diasumsikan mempengaruhi potensi karbon monoksida... (TXT 27)
- (9) The results obtained **in the form of** maps of parameters assumed to influence carbon monoxide... (TXTRANS 27)

2) Functions of LB

Fig. 1 shows the percentage of LBs based on the functions proposed by Hyland (2008). The most frequent LB functions in Indonesian and English abstracts are research-oriented and

text-oriented while the least popular bundles are participant-oriented. This is in line with Wei and Lei (2011) in terms of the rank of the LB functions.

Fig. 1. Percentage of LB functional types



a) Research-Oriented Bundles

Bundles functioning to describe the study such as *penelitian ini adalah* and *this research is* dominate the results at 64% and 82% respectively. Indonesian abstracts have more frequent research-oriented LBs than the English texts. This result is similar to that of Wei and Lei (2011) since, in their study, the bundles functioning to elaborate or describe the research had the highest frequencies.

As an example of this, the writers might wish the reader to pay attention to the elements of the research. A number of the bundles signify the purposes or method as in (10) and (11).

- (10) **The purpose of this research is** clustering emissions sources from several objects by single linkage of hierarchical method in Surakarta (TXTRANS 17)
- (11) **Subjek penelitian ini adalah** individu yang berusia 20 - 25 tahun... (TXT 63)

In addition, similar to Hyland (2008), both in Indonesian and English, the bundle can function to elaborate the procedure of the research as in (12) and (13).

- (12) **Metode penelitian yang digunakan** dalam penelitian ini adalah metode kualitatif. (TXT 21)
- (13) The method that is **used in this research** is qualitative method (TXTRANS 21)

A reason for the frequent research-oriented bundles in both languages might be the need for density of information in the abstracts.

b) Text-oriented Bundles

In contrast to the research-oriented bundles, the text-oriented bundles are more frequent in the English abstracts at 31%

compared to the Indonesian text-oriented LBs, which accounts for 18% of the LB types. The function is mostly resultative, or to signal causative links (Hyland 2008: 14). As an example, the 3-word bundles *in this research* and *dalam penelitian ini* are the most frequent text-oriented bundles. As shown in (14) and (15), the bundles function to refer to the research.

- (14) The methodical elements **in this research** are interpretation, heuristics, and description. (TXTRANS 24)
- (15) Objek formal **dalam penelitian ini** adalah filsafat teknologi Don Ihde (TXT 33)

Another purpose identified is to connect ideas; however, this function seems significantly less common than the resultative function. For example, in Indonesian writers' abstracts, the most common English text-oriented bundle is *based on the* (16) and *according to the*.

- (16) **Based on the** information above, the researcher wants to examine the production system on Processing Unit and Organic Rice Production Tani Mandiri (TXTRANS 6)

The most common Indonesian connectors functioning as text-oriented bundles are *oleh karena itu* and *maka dari itu*, as can be seen in (17) and (18).

- (17) **Oleh karena itu**, popularitas kambing sebagai hewan ternak semakin meningkat sehingga peminat terhadap usaha peternakan kambing semakin meningkat pula. (TXT 1)
- (18) **Maka dari itu** dibutuhkan sebuah sistem informasi yang memudahkan user menangani permasalahan tersebut dengan memberikan informasi yang tepat. (TXT 12)

The appearance of the English/Indonesian bundles above does not necessarily mean that there are no such expressions identified in the other language. Further investigation shows that such bundles are not present in the form of lexical bundles correspondingly. For example, *maka dari itu* is realised as *therefore*. However, it is to be noted that the LBs and their exact correspondences have similar frequencies, showing writers' tendency to use the same words repeatedly.

c) *Participant-oriented Bundles*

In line with Hyland (2008) and Jalilifar et al. (2017), participant-oriented bundles are the least used combinations in the Indonesian and English abstracts. In English, it comprises 5% and there is no 3- or 4-word participant-oriented bundle identified in the Indonesian abstracts.

Most, if not all, of the participant-oriented statements use the bundle *it can be* to signal the writers' attitudes. In Indonesian, there are no 3- or 4-word bundles that can be categorised as participant-oriented. However, a closer examination suggests that its Indonesian counterpart is used in the form of the non-bundle *dapat* although its occurrence is also rare. Excerpts (19) and (20) show how the participant-oriented statements are made in English and Indonesian.

(19) **it can be** known whether the production process is...
(TXTRANS 6)

(20) sehingga **dapat** diketahui apakah proses produksi ..
(TXT 6)

The rather infrequent participant-oriented bundles in both languages and the use of expressions with *can*, which signals certainty, suggest that the writers might want to maintain the objectivity of the texts. As Hyland (2008: 19) states, the avoidance of participant-oriented bundles might have something to do with the mother languages which favour impersonality.

C. *Factors Influencing Lexical Bundle Use*

The first cause might be the students' knowledge of Indonesian and English academic conventions. Based on the data, it could be assumed that the students learned fixed expressions due to the repeated LBs in both languages. Besides that, transfer of L1 knowledge might also contribute to the English LBs as can be seen in the seemingly retained L1 structure in the English LBs, resulting in unnatural use of English word combinations. Secondly, proficiency affects the students' English LBs in the abstracts as some LBs were found to be inaccurate such as the mistaken patterns following *one of*. Next, the differences of the languages themselves could contribute to the LBs of each language. For instance, a 3-word LB in one language can be represented as one word in the other language e.g. *oleh karena itu*, which becomes *therefore*.

D. *Limitation of the Research*

This study has not accounted for the way each lexical bundle is translated. Besides that, it was not ensured whether or not the students wrote the abstracts by themselves.

V. CONCLUSIONS AND RECOMMENDATIONS

The data suggested that the differences between English and Indonesian lexical bundles were not highly significant. The LBs could be regarded similar in terms of the frequencies of the form and functions. However, it could be assumed that the writers of the abstracts were not aware of the academic conventions in the English writing as can be seen in the use of LBs which seem unsuitable for academic research, for example to know the. In addition, the repetitive LBs and inaccuracies of the English LB might signal the students' limited academic vocabulary.

Therefore, it is recommended that teachers of academic writing equip students with the lexical phrases necessary for writing the research abstracts. They can, for example, facilitate students to analyse the common patterns of academic writing LBs in the journal articles and utilise the suitable Academic Word List (AWL) to enrich their vocabulary.

References

- Adel, A. & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes* 31:81–92.
- Anthony, L. (2014) AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net>.
- Biber, D., Conrad, S., & Cortes, V. (2004). 'If you look at...': Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25:371–405.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H.H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: studies in honour of Stig Johansson* (pp. 181-189). Atlanta, GA: Rodopi.
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.
- Baker, M. (1995). Corpora in translation studies: An overview and suggestions for future research. *Target* 7:223-243.
- Chen, Y. H., and 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.

- Dontcheva-Navratilova, O. (2012). Lexical bundles in academic texts by non-native speakers. *Brno Studies in English* 38(2): 37-58.
- Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast* 14(1):58-72.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27: 4-21.
- Jalilifar, A., Ghoreishi, S.M. & Roodband, S.A.E. (2017) Developing an inventory of core lexical bundles in English research articles: a cross-disciplinary corpus-based study. *Journal of World Languages* 1:1-21.
- Lee, C. (2013). Using lexical bundle analysis as discovery tool for corpus-based translation research. *Perspectives: Studies in Translatology* 21(3): 378-395.
- McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory, and practice*. Cambridge: Cambridge University Press.
- McEnery, T, Xiao, R., & Tono, Y. (2006) *Corpus-based language studies: An advanced resource book*. New York: Routledge.
- Ngadiman, A. (2013). The lexical phrases employed in the thesis abstracts of English department students. *Magister Scientiae* 33:93-106.
- Novita, H & Kwary, D.A.(2018). Comparing the use of lexical bundles in Indonesian-English translation by student translators and professional translators. *Journal of Translation and Interpreting* 10(1):53-74.
- Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics* 18(3): 391-417.
- Wei, Y. & Lei, L. (2011). Lexical bundles in the academic writing of Chinese advanced learners. *RELC Journal* 42(2):155-166.

APPENDIX 1
TABLE 2. Most common Indonesian and English LBs in thesis abstracts

RANK	Indonesian 3-word LBs	FREQ	F/10000	Indonesian 4-word LBs	FREQ	F/10000	English 3-word LBs	F	F/10000	English 4-word LBs	F	F/10000
1	penelitian ini adalah	46	29.72	penelitian ini bertujuan untuk	20	12.92	this research is	36	21.03	of this research is	18	10.51
2	dalam penelitian ini	34	21.97	dalam penelitian ini adalah	18	11.63	one of the	25	14.60	in this research is	10	5.84
3	penelitian ini bertujuan	22	14.21	digunakan dalam penelitian ini	12	7.75	in this research	19	11.10	the purpose of this	10	5.84
4	ini bertujuan untuk	21	13.57	yang digunakan dalam penelitian	12	7.75	the result of	18	10.51	is one of the	9	5.26
5	hasil penelitian menunjukkan	15	9.69	ini bertujuan untuk mengetahui	10	6.46	of this study	16	9.35	the result of this	8	4.67
6	yang digunakan dalam	15	9.69	metode yang digunakan dalam	9	5.82	is one of	14	8.18	used in this research	8	4.67
7	data yang digunakan	12	7.75	penelitian ini adalah untuk	8	5.17	the purpose of	13	7.59	of this study is	7	4.09
8	digunakan dalam penelitian	12	7.75	hasil penelitian menunjukkan bahwa	7	4.52	the results of	13	7.59	purpose of this research	7	4.09
9	merupakan salah satu	12	7.75	tujuan penelitian ini adalah	7	4.52	the development of	12	7.01	result of this research	7	4.09
10	yang digunakan adalah	12	7.75	analisis data yang digunakan	5	3.23	to determine the	11	6.43	the results of this	7	4.09
11	hasil penelitian ini	11	7.11	dari penelitian ini adalah	5	3.23	to know the	11	6.43	in the form of	6	3.50
12	penelitian ini dilakukan	11	7.11	penelitian ini menggunakan metode	5	3.23	based on the	10	5.84	the method used in	6	3.50
13	penelitian ini menggunakan	11	7.11	hasil penelitian ini adalah	4	2.58	result of this	10	5.84	this research aims to	6	3.50
14	bertujuan untuk mengetahui	10	6.46	metode penelitian yang digunakan	4	2.58	the level of	9	5.26	this research is to	6	3.50
15	dari penelitian ini	10	6.46	penelitian ini adalah metode	4	2.58	the process of	9	5.26	this study aims to	6	3.50
16	metode yang digunakan	9	5.82	penelitian ini merupakan penelitian	4	2.58	this study is	9	5.26	method used in this	5	2.92
17	ini adalah untuk	8	5.17	pengumpulan data yang digunakan	4	2.58	used in this	9	5.26	results of this study	5	2.92
18	oleh karena itu	8	5.17	tujuan dari penelitian ini	4	2.58	it can be	8	4.67	the result of the	5	2.92
19	penelitian menunjukkan bahwa	8	5.17				significant effect on	8	4.67	the results showed that	5	2.92
20	tujuan penelitian ini	7	4.52				according to the	7	4.09	this research is the	5	2.92

APPENDIX 2
List of top 100 most frequent three-word core bundles in academic writing (Jalilifar et al. 2017)

R	Lexical bundle	F	Rank	Lexical bundle	F
1	as well as	2686	51	the value of	493
2	the number of	1976	52	a variety of	487
3	based on the	1518	53	related to the	484
4	in terms of	1458	54	a series of	471
5	the use of	1432	55	the formation of	466
6	the effect of	1290	56	need to be	458
7	due to the	1208	57	the process of	453
8	the fact that	1089	58	the work of	452
9	according to the	969	59	in relation to	439
10	a number of	965	60	show that the	439
11	the development of	910	61	the form of	439
12	on the other	907	62	it can be	435
13	the result of	893	63	that can be	430
14	the presence of	810	64	as shown in	425
15	the end of	802	65	effect of the	425
16	the role of	780	66	can be used	416
17	the case of	779	67	be able to	412
18	the relationship between	775	68	the present study	412
19	the importance of	765	69	understanding of the	408
20	the other hand	763	70	at the end	407
21	as a result	755	71	the basis of	405
22	lead to the	755	72	a result of	403
23	with respect to	732	73	the amount of	399
24	such as the	708	74	the difference between	395
25	at the same	699	75	the history of	392
26	in the case	651	76	is based on	385
27	the same time	650	77	the university of	385
28	the context of	621	78	value of the	385
29	in addition to	601	79	nature of the	383
30	a set of	600	80	use of the	383
31	in this case	589	81	for example the	382
32	analysis of the	584	82	the quality of	382
33	focusing on the	580	83	the analysis of	377
34	the concept of	576	84	used in the	376
35	well as the	571	85	the majority of	375
36	the study of	550	86	of the studies	374
37	in other words	549	87	be used to	373
38	end of the	541	88	appears to be	359
39	associated with the	540	89	difference between the	358
40	in this study	539	90	the absence of	358
41	in the same	528	91	likely to be	356
42	the level of	527	92	results of the	355
43	the nature of	524	93	the question of	355
44	because of the	521	94	in the following	347
45	in this paper	519	95	refers to the	347
46	suggest that the	518	96	the field of	345
47	depend on the	517	97	in the form	338
48	in the context	509	98	found to be	331
49	the impact of	502	99	a range of	330
50	more likely to	501	100	the existence of	330

List of common four-word core bundles in academic writing (Jalilifar et al. 2017)

Rank	Lexical bundle	Tokens	Rank	Lexical bundle	Tokens
1	<u>at the same time</u>	607	41	in the process of	135
2	<u>in the case of</u>	594	42	the development of the	135
3	<u>as well as the</u>	570	43	in the development of	134
4	<u>on the other hand</u>	559	44	it is difficult to	133
5	on the one hand	544	45	<u>the use of the</u>	133
6	<u>in the context of</u>	496	46	to be able to	131
7	<u>the end of the</u>	448	47	by the fact that	130
8	<u>at the end of</u>	386	48	the context of the	129
9	<u>as a result of</u>	348	49	<u>the results of the</u>	129
10	<u>in the form of</u>	317	50	in contrast to the	124
11	to the fact that	272	51	to the development of	123
12	<u>it is important to</u>	264	52	in the sense that	122
13	<u>on the basis of</u>	255	53	can be found in	119
14	<u>in terms of the</u>	254	54	a result of the	118
15	<u>one of the most</u>	249	55	at the level of	116
16	is one of the	246	56	in other words the	115
17	<u>the nature of the</u>	226	57	<u>with respect to the</u>	114
18	<u>in this case the</u>	213	58	with the exception of	112
19	<u>can be used to</u>	212	59	it is necessary to	111
20	at the university of	209	60	<u>the other hand the</u>	111
21	<u>in the present study</u>	201	61	that there is no	108
22	<u>the fact that the</u>	201	62	<u>the basis of the</u>	107
23	<u>a wide range of</u>	191	63	<u>it is possible that</u>	106
24	<u>the beginning of the</u>	185	64	as one of the	104
25	<u>in addition to the</u>	179	65	it should be noted	104
26	the rest of the	174	66	as well as a	103
27	<u>in the absence of</u>	167	67	the case of the	103
28	<u>the difference between the</u>	163	68	the course of the	95
29	<u>at the beginning of</u>	159	69	<u>the development of a</u>	95
30	that there is a	159	70	the level of the	95
31	in the field of	158	71	the purpose of this	93
32	of the most important	155	72	as a result the	92
33	<u>is based on the</u>	151	73	in the study of	91
34	at the time of	150	74	<u>can be seen in</u>	86
35	the time of the	150	75	for example in the	85
36	the relationship between the	141	76	<u>the presence of the</u>	84
37	it is possible to	140	77	in a number of	82
38	it is clear that	139	78	for the purpose of	79
39	for the first time	137	79	<u>in the next section</u>	74
40	in the same way	137	80	the same time the	70