Lexical Bundles in Applied Linguistics Published Research Articles: A Comparison of L1 and L2 Professionals

Dhinuk Puspita Kirana
Graduate School in ELT Universitas Negeri Malang
& Institut Agama Islam Negeri Ponorogo
dhinuk@iainponorogo.ac.id

Yazid Basthomi
Universitas Negeri Malang
ybasthomi@um.ac.id

Moh.Adnan Latief
Universitas Negeri Malang
mohammad.adnan.fs@um.ac.id

Suharmanto
Universitas Negeri Malang
suharmanto.fs@um.ac.id

Abstract

The current study is a corpus-based investigation of the frequency, structural and functional categories of lexical bundles in the corpora of Applied Linguistics articles written by L1 and L2 English academic professionals. First, four-unit formulaic bundles were extracted from one million word corpus of AL-ENG (Applied Linguistics articles written by L1-English professionals) and one million word corpus of AL-IND (Applied Linguistics articles written by L2-English/Indonesian professionals). The results indicate that the majority of lexical bundles used by L1-English professionals were in noun phrase and prepositional phrase. L2-English (Indonesian) professionals used lexical bundles in verb-phrase about 10% higher than that of L1 professionals. The result of the study is also lending support to the hypothesis which states that both L1 and L2 academic writers may follow the process of advancement and progression from clausal to phrasal styles (Biber et.al., 2011). It is interesting to find out that L1-English professionals seem to use specific bundles in some ways that are distinct from those of L2-English (Indonesian) professionals structurally and functionally (Pan, Reppen, & Biber, 2016:70). The results of this current study indicates that, it is not as simple as a matter of novice or expert, or a matter of L1 versus L2, but it is possible that L1-English obviously has an important and distinctive ability in developing this discourse styles and put them in application. Based on these findings, it is important to note that L2 novice and L2 expert writers need to pay more attention to the transition from clausal to phrasal patterns in an attempt to integrate the information in their academic writing better.

Introduction

In recent years, more studies have been accomplished in the area of corpus linguistics and formulaic sequence (AlHassan & Wood, 2015; Appel & Trofimovich, 2015; Coklin & Schmitt, 2012; Ellis & Maynard, 2008; Gomez, 2015; Hyland, 2008; Hsu, 2014; Jalali, 2009;

Peters & Pauwels, 2015; Staples et al., 2013; Wray, 2013). The study of formulaic pattern (in general) and specific groups of word combination (in particular) has become a broad attention among linguists (Al Hassan & Wood, 2015; Appel & Trofimovich, 2015; Coklin & Schmitt, 2012; Ellis & Maynard, 2008; Gomez, 2015; Hsu, 2014; Peters & Pauwels, 2015; Staples et al., 2013; Wray, 2013).

Among diverse categories of formulaic sequences, lexical bundles, also known as 'clusters' or 'chunks' or 'extended collocation which occur more regularly than expected by coincidence, allowing to form the implications and providing to the sense of attachments in the texts (Hyland, 2008). Decades of research on phraseology has determined the significance of multi-word collocation or collocation in the natural production of English (Biber et al., 1999). Experts believed that language is formulaic in nature and it was found out that the various types of formulaic sequences constitute 21% of the written discourse (Biber, D., & Conrad, S.,1999).

There are studies investigating the functional contributions to the discourse coherence, written or spoken in different kinds of texts (Cortes, 2004; Biber et al., 2004; Herbel-Eisenmann & Wagner, 2010; Hyland, 2008, 2012; Jablonkai, 2010; Jalali et al., 2014a; Nesi & Basturkmen, 2006). Lexical bundles have become a crucial component of academic discourse (Biber et al., 1999). There are also research concerning on academic writing production in research articles (Cortes, 2004); writing production in intensive history class (Cortes, 2007); applied linguistics articles (Kazemi et al., 2014); and research articles on biology (Sánchez, 2014).

The studies investigating the different uses of formulaic language written by native and non-native writers have been great interests in the field of corpus linguistics and English for Academic Purposes (Ädel, A., & Erman, B., 2012). A study conducted by Chen and Baker revealed that lexical bundles in published academic writing was found to be in a wide range use while L2 student writing showed the smallest range (Chen and Baker, 2010). It also in line with the result of the study conducted by Dontceva, (2012) which has evidenced that the frequency use of lexical bundles by Czech students (L2) was lower than typical expert academic discourse. In addition, several most used recurrent expressions in published texts were rarely used by both L1 and L2 students while some lexical bundles which native academics rarely used were used by

L2 student writers excessively (Baker, 2010). It was assumed to be due to of insufficient level of writing skills of L2 writers.

Furthermore, it was found that, as the level of study increased, students tended to use a larger quantity and variety of lexical bundles (Qin, 2014). On the other hand, result of a study involving EFL Junior and Senior Chinese students indicated that although the senior students tended to use a wider range and variety of lexical bundles than the junior ones, it does not mean that the seniors have used them more precisely than the junior ones (Huang, 2015). Hence, it can be said that although some senior students use more bundles, it does not guarantee that they used them accurately.

The frequent use of lexical bundles can be used as the signal of competent language use in particular register. Heng et al., (2014) states that the more competent students are more naturally and sometimes unintentionally utilized specific bundles in order to send definite discourse functions within their arguments. On the other hand, according to the other previous studies, student writers rarely use lexical bundles within their academic manuscripts and there are misuses of language due to their lack of awareness in choosing the correct and more natural formulaic patterns. The other study conducted by Cortes (2004) indicated that student writers rarely used lexical bundles in their writing. Furthermore, even when they use certain lexical bundles, some of them misuse them inappropriately or did not match with the uses of lexical bundles employed by expert authors.

There is a disagreement among scholars on whether there are distinctions between L1-English writers and L2-English writers in using lexical bundles (Pan, F., Reppen, R., & Biber, D., 2016). Therefore, it is crucial to investigate whether discrepancies between L1 and L2 continue to happened to professionals/ experts or whether those differences disappear at the professional/ expert levels. The data we used to answer the questions were academic research articles on Applied Linguistics discipline, written by L1-English writers and L2-English (Indonesian) writers writing in English, published in reputable journals.

Applied linguistics discipline was chosen with an assumption that the professional writers who write in the area of applied linguistics are professionals who are supposedly very aware of the language aspects or language features including formulaic expressions which are

expressed through the use of lexical bundles. Applied Linguistics uses language-related research in a wide variety of fields (e.g. language acquisition, language teaching, literacy, gender studies, language policy, speech therapy, discourse analysis, censorship, workplace communication, media studies, translation, lexicography, forensic linguistics). Applied linguistics is the usage of the knowledge about the nature of language achieved by linguistic research for the improvement of the efficiency of some practical task in which language is a central component (Corder, 1974, p. 24).

There has been much research focused on identifying and how lexical bundles are used across disciplines and by student and expert writers, while this study is focusing on native and non-native expert writers. The outcomes of the research are significant that they offer insights into the distinctive character of scientific writing by revealing the main concerns of expert writers and the ways in which they construct their arguments and pursue their purposes.

The comparison between L1-English and L2-English (Indonesian) professional writers underscores the differences between the native and non native expert writers and later on can pinpoint in give contributions and theoretical significance to phraseology research. This study is significant for language students to come to a better understanding of the word-combination, lexical bundles. This study is also be beneficial to the EFL students in employing lexical bundles to establish their opinion and arguments to better communicate their ideas to the readers. Moreover, this study is helpful to the students since the study of lexical bundles offers insights into elements of language in use, maintaining the sensitive connection between variety bundles uses in different contexts. The findings of the research are useful for learner writers in obtaining naturalness in English academic writing.

Research Questions

1. What structural distinctions of lexical bundles exist in published Applied Linguistics research articles written by L1-English and L2-English (Indonesian) professionals?

2. What functions of lexical bundles exist in published Applied Linguistics research articles written by L1-English and L2-English (Indonesian) professionals?

Research Design

The study employed corpus based approach. The two corpora used in this study were Applied Linguistics (AL) research articles written by L1-English writers (ENG) and by L2-English (Indonesian) writers (IND). The AL-ENG corpus research articles were collected from the best scholarly research journals on Applied Linguistics: *Applied Linguistics, Cognition, College English, Language Learning and Technology, Language Learning, TESOL Quarterly, Journal of Second Language Writing,* and *System*. The AL-IND corpus was compiled from selected articles appeared in major Applied Linguistics research journals published in Indonesia: *TEFLIN (Teaching of English as a Foreign Language in Indonesia), IJAL (Indonesian Journal of Applied Linguistics)* and *CONAPLIN*.

In order to ascertain the first language of the authors, method proposed by Wood (2001) was applied in this study. This study used the definition of 'L1-English' writers as previously mentioned by Wood (2001) who operationally defined as any author affiliated with any institution in a country where English is used as the first medium for communication. The writer's name should also have the first name and the last name which can be representative into native to English-speaking countries Wood (2001). The author whose name is questionable is excluded although her/his affiliation is universities or institution in the English-speaking countries. The writers who are considered having native English names, but are affiliated in institutions outside English-speaking countries were excluded. L2-English (Indonesian)' writers were operationally defined as any authors who have names which can be considered as native Indonesian names and affiliated with an institution in Indonesia. The Indonesian writers who are affiliated in institutions which are not in Indonesia were excluded.

The current study adopted the definition of formulaic language from Biber et al. (1999), as '*lexical bundles*'. These word-sequences were identified as *lexical bundles* by a computer program when they occur minimum 10 times in a million words in a corpus and found in minimum 5 distinct manuscript in the corpus. These empirically identified bundles proved a statistical trend to take place together in natural discourse.

The lexical bundle list was collected from the raw data (corpus of AL-END and corpus of AL-IND). The process was made possible by employing ABBYYFineReader 8.0 program and

AntConc 3.4.3 software (Anthony, 2007). The 4-word-exical bundles were listed using N-gram programs and then categorized according to their structural categories proposed by Biber et.al. (1999; section 1.2) and grouped into three main categories: NP-based, PP-based, and VP-based, similar to adapted classification by Chen and Baker (2010: 35). Next, the lexical bundles were analyzed in their context of situation by concordance analysis and file view from AntConc3.4.3 software. After that, they were categorized according to their functional categories of bundles by Hyland (2008a, pp.13-14; 2008b, pp.49).

In the current study, the researcher focused on investigating only four-word bundles, since four-word bundles are in a greater extent than five-word bundles and provide a larger range of functions. Moreover, according to Hyland (2008b), 4-word bundles are much more found in the corpora rather than 5-word bundles and they have obvious variations of structures and functions than 3-word bundles. The researcher set the 20 recurring bundles in a million words and the dispersions minimal 5 different texts. The researcher also utilized the concordancer in *Anconc 3.4.3 software* in exploring the current context of situations where the bundles had been employed.

The next analysis is the qualitative analysis. It is concerned with the relationship between text and the bundles. The data were analyzed qualitatively starting from the forms/structures and the functions of the bundles. The structural forms are categorized into 3 main categories: noun-based, prepositioned-based, and verb-based bundles (Biber et al., 1999).

The next step, the lexical bundles were analyzed according to the 3 main categories of functions of bundles. The functional classifications established by Biber (Biber et al., 2004; Biber, 2006) were then developed extensively, specified more on lexical bundles on academic writing by Hyland (2008a, 2008b). Thus, Hyland's functional taxonomy classifies three broad categories: research, text and participants. These classifications represent the detailed focus on research writing. They are: (1) bundles related to research, (2) bundles related to text, and (3) bundles related to participant. See details in Table 1.2 Functional categories of lexical bundles (Hyland, 2008a, p.13-14, 2008b, p.49) in the theoretical framework.

The concordance program was used again in analyzing the lexical bundles in their contexts and determining the specific functions that they perform. The approach involved

6

analyzing all instances of every lexical bundle on the list in its context of use in order to reveal the discourse function of each lexical bundle.

Findings and Discussions

From 1.023.069 tokens, the finding of the research shows that there are 169 types of 4-word lexical bundles and 5708 tokens. On the other hand, from 1.004.409 tokens from the AL-IND Corpus, the result shows that there are 224 types of 4-word lexical bundles or 8803 tokens. As can be seen in Figure 1, L1-English professionals use fewer bundles (169) compared to L2-English (Indonesian) professionals (224). The result of the study shows that the wider number of bundles utilized in the corpus AL-IND is the preceding sign which explain that L2-English (Indonesian) professionals rely more on lexical bundles in a considerably above average than the L1-English professionals.

Overall, the L2-English (Indonesian) professionals used lexical bundles with a greater frequency than did the L1-English professionals, 224 and 169 respectively, with 71 bundles used in both groups (42% of lexical bundles of L-1 English professionals and 31% of lexical bundles of L2-English (Indonesian) professionals). In other words, the total bundles employed in the corpus of AL-IND were approximately 25% higher in number than bundles exist in the corpus AL-ENG corpus. The total application of lexical bundles can be seen from the total tokens of the total frequency per million words used in each of the corpus. As mentioned earlier in the previous section, there were 5781 tokens of bundles used in AL-ENG corpus and 8803 total tokens used in AL-IND corpus. It can be concluded that there were 34% more bundles which were applied in the corpus AL-IND than bundles applied in the corpus AL-ENG.

Therefore, the fact that L2-English (Indonesian) professionals tended to use a larger number of distinctive bundles and also the total frequency than L1-English professionals was the preliminary indication that L2-English (Indonesian) professionals relied on lexical bundles to a wider extent than the L1-English professionals did. Despite some of the previous studies argue that non-native speakers of English use fewer lexical bundles (Erman, 2009; Howarth, 1998) and less various (Granger, 1998), the findings of the present study are vice versa. Thus, the findings

are consistent with several studies which revealed that non-native speakers use a broad range of lexical bundles (Hyland, 2008b; Öztürk, 2004; Pérez-Llantada, 2014; Pan, Reppen, & Biber, 2016).

The most frequently used bundles

The most frequently used bundles in AL-ENG corpus was *'the extent to which'* which appeared 144 times in 54 different texts in the corpus. Furthermore, the second mostly used bundles in the corpus AL-ENG was also the most frequently used bundles in AL-IND corpus *'on the other hand'*, appeared 142 times in 64 out of 135 different texts within AL-ENG corpus, and also appeared 191 times in 105 out of 195 different texts in AL-IND corpus. In addition, the second most frequently used bundles in AL-IND was *'in the form of'* which appeared 179 times in 83 different texts within the corpus.

Shared bundles in both corpora

The findings showed that there were 71 different types of lexical bundles used in both corpora (shared bundles), for instance: *on the other hand, the end of the, in terms of the, at the same time, as well as the,* and *in the context of* were some instances. In terms of variety, the finding showed that 42% of lexical bundles used by L-1 English professionals were also used by L2-English (Indonesian) professionals, while only 31% of lexical bundles used by L2-English (Indonesian) professionals were also by L1-English professionals.

There were 3013 tokens of shared bundles in AL-IND corpus and 3894 tokens of shared bundles in AL-ENG corpus. Therefore, it means that the shared bundles in AL-IND corpus were approximately 29% than those bundles shared in AL-ENG corpus. In terms of the overall frequencies of these shared bundles, it was found that the overall frequency of these shared bundles in AL-ENG corpus was 3013 tokens (in one million) and in AL-IND corpus was 3894 tokens. Therefore, it means that the overall use of shared bundles in AL-IND corpus were approximately 29% more than that in AL-ENG corpus.

The lexical bundles from the AL-ENG were used by AL-IND authors, but with lower frequency. There were a wider range of lexical bundles which were utilized by L1-English

professional which were not used by L2-English (Indonesian) professionals. It can be estimated that 58% of the lexical bundles in AL-ENG were not used in AL-IND. The findings showed that there were 98 different types of bundles in AL-ENG which were not used in AL-IND, for instance: *the extent to which, over the course of, a wide range of, in the case of, the ways in which, the degree to which, with respect to the,* and others.

On the other hand, there were 153 bundles which were specific to AL-IND. Many of them were the repetition of terminology in applied linguistics, for instance: *the teaching and learning, teaching and learning process, as a foreign language, English as a foreign, the teaching of English, of English as a, of the teaching and learning, in the Indonesian context, Ministry of National Education, etc.*

Here are the examples of some specific bundles in AL-IND, taken from the articles using the concordancer.

(1)   In preparing their students for ENE, the teachers focused their teaching on listening and reading aligning the teaching with the SKL UN (Standar Kompetensi Lulusan or the table of specification of ENE) included in the policy of the **Ministry of National Education** Number 75/1995 on national examinations. (AL-IND, TEFLIN 015, 2015)

(2)   It is expected that this study which implemented R2L pedagogy **in the Indonesian context** will contribute to English language teaching in EFL contexts.(AL-IND, IJAL 060, 2017)

(3)   Overall, this derives from a personal account of teaching Advanced Applied Linguistics to doctoral students at the School of Graduate Studies (PPs), **State University of Malang** (Universitas Negeri Malang) during the second semester of the 2011/12 academic year. (AL-IND, TEFLIN 027, 2014)

These findings can be compared with other studies to demonstrate that these lexical bundles are the characteristics of the disciplinary discourse of the Applied Linguistics academic field. Table 4 compares the top 50 AL-ENG and AL-IND four-word bundles with the top 50 most frequent bundles in Hyland's research (2008) from almost 800,000 word corpus of research articles, PhD dissertation and MA/MSc theses in Applied Linguistics. The bundles in bold are among 50 most frequent lexical bundles in Hyland's study which also appeared in both AL-ENG and AL-IND corpora, while italicized bundles are those bundles which existed in Hyland's study and AL-ENG corpora.

There were more than half of the bundles used in Hyland's study were also used in AL-ENG and AL-IND corpora: 27 bundle types (54%) bundles out of 50 most frequently used bundles in Hyland's study. There is the consistency between Hyland's 2008 list of lexical

bundles on applied linguistics and the list of lexical bundles in AL-ENG and AL-IND corpora. This can be the indication of the validity of both findings of the studies which demonstrate that those lexical bundles are the typical characteristics of the disciplinary discourse of the academic field: Applied Linguistics. Those lexical bundles are useful to the learners who want to comprehend and utilize them in research-focused texts in the specific field of study.

Comparison of the distribution of structural categories

The similarities and distinctions between both corpora were analyzed by contrasting and comparing the lexical bundles' main categories including each subcategories and their respective percentage for their use, including the types and total frequency per million words. The researcher employed the structural taxonomy established by Biber et. al., (1999: 1015-1024). To go deep into the analysis, the lexical bundles were grouped into three primary classifications: Noun phrase-based (NP-based), Prepositional-phrase based (PP-based), and Verb-phrase based (VP-based). The three main categories were later analyzed in more detail types. NP-based bundles include noun phrases; PP-based bundles include prepositional phrases; and VP-based bundles, which related to bundles with a verb element (Chen & Baker, 2010). The primary distinction is between phrasal bundles and clausal bundles. The phrasal bundles are normally formed by noun phrases and prepositional phrases, while clausal bundles generally include simple verb phrases and also bundles which include a main clause (Biber et al., 1999).

Comparison of the Distribution of structural subcategories

The AL-ENG and AL-IND corpora were categorized into three main categories: NP-based, PP-based, and VP-based, similar to the previous study by Chen and Baker (2010). NP-based bundles include the noun phrases, PP-based bundles include prepositional phrases, and VP-based bundles refer to the combination of words with verb component (Chen & Baker, 2010).

Noun-phrase based

The findings showed that the noun phrase + of-phrase fragment was the second mostly used structure in the AL-ENG and AL-IND, 22% and 21% respectively. Together with noun

phrases with other post-modifier fragments and other types of noun phrases, they comprised 33% of all bundles in AL-ENG corpus and 31% of all corpus in AL-IND.

This result coincides with recent findings which show that NP-based bundles are widely used in academic writing (Biber et al., 1999; Hyland, 2008, and Byrd & Coxhead, 2010). In addition, these results also support the view of academic writing as being "noun-centric" (Swales, 2008: 5).

Prepositional-phrase based

Among the three main structural categories, the frequently used lexical bundles rely on repetition of PP-based bundles, with the total number of 47.3% of all lexical bundles used in AL-ENG corpus and 40.2% of all bundles used in Al-IND corpus. In the case of structural comparison, both corpora tended to rely on the same structural bundles: noun phrases and prepositional phrases.

The result of the study support the finding of the previous studies which stated that in contrast with classroom teaching and conversation, the most noticeable quality of bundles used in academic writing were mostly phrasal rather than clausal (Biber et.al., 1999; Cortes, 2002, 2004).

Verb-phrase based

Verb-based clausal structures represent 17.7% of all AL-ENG corpus and approximately 25.5% of all AL-IND corpus. The majority of verb structures are composed of passive verb followed by prepositional phrase 13% from all VP-based bundles in AL-ENG corpus and 10% of all VP-based in Al-IND corpus.

The result of the analysis also found that L1-English professionals mainly use phrasal bundles: noun phrases and prepositional phrases, with the total 80.5 % of all the types of bundles used and approximately 83% of the tokens or the total frequency per million.

The VP-based bundles found in corpus LL-IND was around 8% higher than those found in corpus LL-ENG (25,4%) and (17,7%). On the other hand, the phrasal bundles found in AL-IND corpus were also consistent with the bundles used in AL-ENG corpus, about the similar

percentage, with the total 80%. Surprisingly, L2-English (Indonesian) academic professionals also relied more on NP-based bundles (30%) and PP-based bundles (40%).

These findings were consistent with the result of the previous studies conducted by Biber & Conrad, 1999; Biber et al., 1999, 2004; Biber & Gray, 2010; Byrd & Coxhead, 2010 which show that the majority of lexical bundles used in academic writing are phrasal bundles (NP-based and PP-based bundles) rather than clausal.

Previous studies have shown that the more attentive combination of particular sequences in academic compositions was needed to later sift from clausal style to phrasal or prepositional. The correct use of phrasal features in writing academic prose is the indication of its high focus of the information mentioned (Pan, Reppen & Biber, 2016).

Comparing the distribution of functional categories

The three primary functions of bundles were matched to Halliday's (1994) tripartite of language metafunctions. The purpose of examining the lexical bundles was to investigate to what extent did the L1-English professionals and L2-English (Indonesian) professionals use lexical bundles for similar or different functions. Bundles related to research give a more ideas in explaining activities, experiences, and practices in the actual world; bundles related to text present textual functions in organizing and connecting various parts of discourse; and bundles related to participants involve a more interpersonal role by developing interactional communication between writers and readers (Hyland, 2008a: 13-14, 2008b: 49).

The finding showed that that among the three sub-categories, the most frequently used functional sub-type was text-oriented bundles. Both L1-English native professionals and L2-English (Indonesian) professionals used the greatest number of text-oriented bundles with similar frequency, with almost half of all bundles used in each corpus: 49.70% in AL-ENG corpus and 48.66% in AL-IND corpus. The finding of this study consistent with the previous study which says that the applied linguistics studies corpora were dominated by text-oriented functions (Hyland, 2008a: 16). AL-ENG writers prefer to use text-oriented bundles which "help writers to structure their activities and experiences of the real world" (Hyland, 2008b, p.49).

The second most used function was research-oriented; 34.32% in AL-ENG corpus and 36.61% in AL-IND corpus. Unlike the text-oriented bundles, L2-English professionals used

higher percentage of research-oriented bundles, approximately 2% higher. Both corpora (AL-ENG and AL-IND) used the participant-oriented in the similar percentage; 15.98% and 14.73% respectively.

L1-English professionals and L2-English (Indonesian) professionals tended to use participant-oriented bundles infrequently compared to the other 2 main functional categories. Both AL-ENG and AL-IND writers showed the least preference in using participant oriented bundles, 16% and 15% respectively. This finding is relevant with other study which found that 51.3% of bundles were research-oriented, 42.4% were text-oriented, and the least were participant oriented bundles with only 6.4% of all bundles used (Salazar, 2010).

Comparing the distribution of functional subcategories

Within the functional category, there are three subtypes: research-oriented bundles (bundles related to research), text-oriented bundles (bundles related to text) and participant-oriented bundles (bundles related to participants). Bundles related to research make the researcher easier to organize their actions and practical activities in the real world, which is consisted of 5 subcategories: location, procedure, quantification, description, and topic. The text-oriented bundles concern in arranging elements within the whole text systematically as ideas or reasoning: transition signals, resultative signals, structuring signals and framing signals. The participant-oriented bundles focus on the writers or readers of the text: stance features and engagement features (Hyland, 2005).

Text-oriented bundles

It has been stated earlier in the previous section that L2-English (Indonesian) professionals used around 33% bundle types more than that of L1-English professionals. To begin with, the researcher firstly discussed the research-oriented bundle sub-category: location, which mark the period of time and place or location. The total frequencies of location-bundles were similar between the two corpora, where L1-English professionals used about the same percentage with L2-English (Indonesian) professionals. In terms of the total frequencies of location-bundles, the result showed that there were about 3.5% and 2.68% of all bundles used in

each corpus respectively. *At the end of,* and *at the beginning of*, were the two shared bundles used in both corpora.

In terms of procedure bundles, L2-English (Indonesian) professionals tended to use procedure-bundles in a wider range of types than that of L1-English professionals, approximately 26.52% and 9.45% of all research-oriented categories respectively. Procedure bundles in AL-IND corpus was the most frequent bundles among the four other sub-categories in research-oriented. As has been mentioned previously, procedure bundles were used to explain experimental procedures and scientific phenomena. Procedure bundles are mostly past-tense passive structures that describe research activities and experimental techniques, for instance: *can be used to, used in this study, participants were asked to, the data were collected, used by the students, the students were asked to, is used as a, be used as a,* and *strategies used by.*

Quantification-bundles encode the quantity, amount, degree, number, variety and nominal entities. These bundles were one of the widest ranges of bundles used in both AL-ENG corpus. The quantification bundles used in AL-ENG corpus was much higher than that of the AL-IND corpus. L1-English professionals tended to use quantification bundles in a wider range of types and frequency than L2-English (Indonesian) professionals. L1-English professional used quantification for more than 50% higher than that of L1-English professionals, 11.83% and 6.70% respectively. *A wide range of, the total number of,* and *the number of words* were the three most frequent bundles used in AL-ENG corpus, while *is one of the, as one of the* and *the majority of* were the three most frequent bundles used in AL-IND corpus.

The distribution of description-bundles was about the same amount within AL-ENG and AL-IND corpora, with the total of 5.33% and 4.46% respectively. L1-English professionals used description-bundles in a quite similar percentage with L2-English (Indonesian) professionals, both in types and tokens. In Applied linguistics, the bundles are tied to be used in describing studies, explaining methodology, reporting results and specific rhetorical moves in the field of study. There were 4 types of description-bundles which were shared in both corpora: *the meaning of the, the nature of the, the quality of the* and *at the level of.*

The L2-English (Indonesian) professionals tended to use more topic bundles than that of the L1-English professionals with the percentage of 6.25% and 4.14% respectively. The wider range of the topic bundles in the AL-IND corpus could be explained by the repetitive use of these bundles in AL-IND, for example: *English as a foreign, the teaching of English, of English as a, native speakers of English, in the Indonesian context, a foreign language EFL, of English language teaching, State University of Malang, English as a second, by the Ministry of, English as an International,* and *Ministry of National Education.* On the other hand, in AL-ENG corpus, there was repetitiveness use of these bundles: *the L1 and L2, of L1 and L2, English as a second, English as a foreign* and *independent samples t-test.* The examples of description-bundles used in context were listed below.

(4)   It is expected that this study which implemented R2L pedagogy *in theIndonesian context* will contribute to English language teaching in EFL contexts. (AL-IND, IJAL 060, 2017)

(5)   The population of this study was the fourth semester students who were officially registered at English Department *State University of Malang.* (AL-IND, IJAL 001, 2016)

(6)   All categories that were significant between early arrival Generation 1.5 and L2 texts were also significant in the same direction between *the L1 and L2* texts (9 out of 11 variables). (AL-ENG, Journal of Second Language Writing 009, 2017)

(7)   Initial differences between students in abroad groups 1 and 2 were assessed using *independent samples t-tests.* (AL-ENG, System 001, 2017)

Research-oriented bundles

Turning in to the text-oriented bundles, the results showed that text-oriented strings accounting for almost 50% of all bundles used in each corpus, approximately 50% of all bundles types used in AL-ENG corpus, and about 49% of all bundles used in AL-IND corpus. These findings are lending support to previous study which stated that bundles in research articles serve a textual function to a large extent, more discursive function of marking the relationship between prior and coming discourse (text-oriented bundles) (Hyland, 2008a, 2008b).

In AL-ENG corpus, the second most frequently used bundles after framing signals were transition signals (21%), and then followed by resultative signals (19%), and structuring signals were the least (15%). On the other hand, the second most frequently used bundles after framing

signals in AL-IND were resultative signals (25%), and then followed by structuring signals (18%), and Transition signals were the least (16%).

The results of the analysis showed that L1-English professionals tended to use more transition-signals than that of L2-English (Indonesian) professionals, accounted for 10.65% and 8.4% of all text-oriented bundles in each corpus. Surprisingly, it can be seen from the Table that the total number of types were at the same number, i.e. 18 different types of bundles. Among the 18 types of bundles, there were almost 50% of them were the shared bundles (8 types): *on the other hand, at the same time, as well as the, on the one hand, in other words the, in addition to the, in a way that,* and *as well as to.* L2-English (Indonesian) professionals tended to use resultative-signals in a wider range of types than that of L1-English professionals, with the total number of 12.05% and 9.45% of all bundles. There were 7 bundles in these types shared across corpora: *as a result of, the results of the, in relation to the, the results of this, results of this study, that there is a,* and *as part of the.*

L2-English professionals tended to use bundles in the form of transition-signals more than that of Ll-Indonesian professionals, accounted for about 10.65% and 8.04% of all bundles in AL-ENG and AL-IND corpora. There were only 4 shared bundles among them, they were: *in the present study, of the present study, as shown in Table,* and *in this study were.* Among the 4 subcategories in text-oriented bundles, Framing was the most frequently used bundles in both AL-ENG corpus and AL-IND corpus, with the percentage of 21.89% and 19.64% of all text-oriented categories. This finding was slightly smaller than the previous study conducted by Hyland, where around 50% of text-oriented bundles in social science texts, including applied linguistics, worked to frame the writers' arguments by emphasizing connections, specifying cases and pointing to limitations (Hyland, 2008a: 16).

Although there were bundles which were unique to each of the corpus, there were about 54% bundles in framing signals which were shared among the two corpora. For instance, the bundle '*in the field of*' was the most frequently used bundle in AL-ENG corpus, used as many as 133 times in 29 texts out of 135 texts (21.5%). On the other hand, it was rarely used in AL-IND corpus, only 35 times in 22 texts out of 195 texts (9.8%). There were 20 shared lexical bundles in

this function. For instance: *in the field of, in the form of, the end of the, in the context of the, in terms of the, in the field of* and others.

Participant-oriented bundles

Both L1-English professionals and L2-English (Indonesian) professionals tended to use a quite similar number of participant-oriented bundles, with the total 16% of all bundle types used in AL-ENG corpus, and 15% of all bundle types used in AL-IND corpus. L1-English professionals used more stance features than L2-English (Indonesian) professionals. There were about 67% of stance bundles and 33% engagement features of participant-oriented bundles in AL-ENG corpus. On the other hand, surprisingly, L2-English (Indonesian) professionals tended to use engagement features more than that of L1-English professionals. About 58% stance bundles and 42% engagement bundles found in Al-IND corpus, in terms of participant-oriented bundles.

Although both corpora had about the same number of lexical bundles in this function (18 and 19 different types) and used about the same total frequency (583 and 615), but they only shared 2 types of bundles: *to be able to* and *an important role in*. There were many bundles unique to L1-English professionals as well as unique to L2-English (Indonesian) professionals. Furthermore, L2-English (Indonesian) professionals tended to use the engagement bundles more than that of L1-English professionals, not only the types but also the total frequency. The total frequency used by L1-Indoensian professionals was almost 50% higher than L1-English professionals did (527 tokens compared to 282 tokens). There were only three shared bundles found in both corpora in the engagement function: *it is important to, can be seen in,* and *as can be seen.* The results of the study support the previous studies, in which participant-oriented bundles were used to express different kind of stance meaning and encode engagement features (Hyland, 2008a, 2008b) are the smallest number of bundles used compared to the other two main categories (Biber, 2006).

Conclusion and Suggestions

The findings of the present study show that the majority of lexical bundles used by L1-English professionals were in noun phrase and prepositional phrase. Although L2-English (Indonesian) professionals also used lexical bundles in noun phrase and prepositional phrase, but

it was 10% lower than that of L1-English professionals. On the other hand, L2-English (Indonesian) professionals also used lexical bundles in verb-phrase (especially in passive verbs structures) about 10% higher than that of L1 professionals. There were structural distinctions between the two corpora, although they were all experts in their academic discipline and are already at their professional level.

L1-English and L2-English (Indonesian) professionals did not have many distinctions in terms of numbers of the functional distribution of bundle types and tokens or total frequency per million words. It was, however, L2-English (Indonesian) professionals used fewer research-oriented bundles than L1-English professionals. It is important to note that L1-English professionals seem to use specific bundles in some ways that are distinct from those of L2-English (Indonesian) professionals structurally and functionally. This result lends support to Pan, Reppen, & Biber, (2016: 70).

Based on these findings, it is important to note that L2 novice and L2 expert writers need to pay more attention to the transition from clausal to phrasal patterns in an attempt to integrate the information in their academic writing better. The results of this current study indicates that, it was not as simple as a matter of novice or expert, or a matter of L1 versus L2, but it was possible that L1-English obviously has an important and distinctive ability in developing this discourse styles and put them in application.

This current study had several limitations to consider. First, this study only focused on one discipline, Applied Linguistics. Second, this study only compared L1-English and L2-English (Indonesian) professional writers writing research articles in published journals. Future studies should be able to compare the research articles written by L1-English and L2-English (Indonesian) professionals in different disciplines, different types of academic writing, or within the sections of research writing to better understand the variety of lexical bundles employed by L1 and L2 academic professionals.