

CHAPTER I

INTRODUCTION

This chapter presents the background of the study which covers the reasons why this study needs to be carried out. Besides, it also presents the state of the problems which are the focus of the study, the objectives – what the study attempts to achieve, as well as the significant roles of the present study in the development of corpus linguistics, especially in the analysis of lexical bundles.

1.1 Background of the Study

The study of groups of words or sequences of words has become the focus of numerous linguists and scholars instead of individual words for decades. It is because of recent theories (e.g., Biber et al., 1999; Hyland, 2008a, 2008b, Salazar, 2014), as well as empirical evidence (e.g., Kwary et al., 2017; Damchevska, 2019; Ren, 2021; Yin & Li, 2021) have shown that multi-word sequences are the essential part of the language (Yuliawati et al., 2021). Besides, it has been argued that the control of the multi-word expressions also referred to as clusters, chunks or bundles is a crucial element of fluent linguistic production and has an important role as essential building blocks of discourse (Biber et al., 2004; Hyland, 2008a, 2008b).

Despite examining the same object of study, some scholars have their terms for describing the multi-word expressions or lexical bundles, for instance "cluster" (Scott, 1997), "n-grams" (Forchini & Murphy, 2008), and "lexical phrases" (Li & Schmitt, 2009). Additionally, Simpson-Vlach & Ellis (2010) preferred to use "academic formulas", and Pérez-Llantada (2014) use the term "formulaic language". Lexical bundles are frequently occurred multi-word expressions in a

particular register and are commonly unseparated in a natural discourse which helps to produce meanings, and provide support to our coherence sense and uniqueness in a text (Biber et al., 1999; Biber, 2006; Hyland & Jiang, 2018).

Some studies of lexical bundles are conducted particularly in academic discourse, for instance, Nesi & Basturkmen (2006), and Liu & Chen (2020) focused on examining lexical bundles in academic lectures. As for academic writings, lexical bundles have been carried out in various genres/types of text, for instance in high school-graduate and university textbooks (Biber et al., 2004; Alasmay, 2019; Alzahrani, 2020); graduate students' thesis (Wachidah et al., 2020); dissertation/Ph.D. abstracts (Lu & Deng, 2019; Li et al., 2020); and argumentative essays (Bychkovska & Lee, 2017; Shin, 2018, 2020; Appel & Murray, 2020; Lee et al., 2020).

Another type of academic writing that has been commonly used to examine lexical bundles is academic journal articles, especially research articles (RAs). Some scholars focused on one particular discipline, for instance, Psychology (Esfandiari & Barbary, 2017), Economics (Damchevska, 2019), and Language/Linguistics (Amalia et al., 2021). On the other hand, other scholars involved two disciplines, such as Ren (2021) in Applied Linguistics and Pharmacy; four disciplines, such as Hyland (2008b) in Electrical Engineering, Biology, Business Studies and Applied Linguistics; Kwary et al. (2017) in multiple major or wide disciplines, such as health sciences, life sciences, physical sciences, and social sciences; Hyland & Jiang (2018) in Applied Linguistics, Biology, Engineering and Sociology; Yin & Li (2021) in Finance, Accounting, Biology, and Applied

Linguistics; and six disciplines, such as Omidian et al. (2018) in Mechanical Engineering, Physics, Biology, Sociology, Marketing, and Applied Linguistics; and Budiwiyanto & Suhardijanto (2020) in Medical Science, Nursing Science, Chemistry, Computers, Philosophy, and Legal Studies.

Research articles (RA) may be the most frequently used genre by most researchers in researching academic discourse, especially lexical bundles compared to other types of academic text. Hyland (2009) argues that research articles (RA) are an outstanding academic genre due to some reasons. First, the value that is bound by the scientific institution to the processes of peer review as a control mechanism for converting beliefs into knowledge. Second, the prestige related to this genre reorganizes the process of thought and research it describes to create a discourse for the sake of the production of scientific fact.

All studies that have been previously mentioned certainly give general comprehension and provide numerous insights and perspectives to the author on how lexical bundles should be examined. However, most of these studies, especially those that are concerned with the analysis of lexical bundles of research articles in multiple disciplines, have only been focused on the structural forms – the classification of their internal structure (Biber et al., 1999) and the functional classifications – a taxonomy based on their discourse functions (Biber et al., 2004; Biber, 2006; Hyland, 2008a, 2008b).

Furthermore, as what has been asserted by Shin (2018), the functional classifications analysis in various studies of lexical bundles has been conducted without drawing the correlation between particular types of contexts (e.g., the

semantic domains) and particular bundles. Therefore, as an attempt to fill in the gap, the current study not only investigates the structural forms and functional classifications of lexical bundles but also extends the functional classifications analysis by focusing on the semantic domains, particularly semantic preference. In this case, the semantic preference is determined based on the shared bundle(s), that is, bundle(s) that are commonly found in the four disciplines.

The journal articles of the current study are collected from *PLOS ONE* journal. There are some reasons why *PLOS ONE* journal is chosen as the source of journal articles that are used as the corpus data in the present study. First, based on the information of the journal, *PLOS ONE* is an open-access journal, meaning that it is freely accessible to everyone without any restriction (PLOS ONE, 2022). It is an essential criterion since researchers who want to conduct a corpus study often have a problem in building their corpus due to copyright issues.

Second, it is a journal that publishes multidisciplinary, and even interdisciplinary articles based on more than 200 subject areas across science, engineering, medicine, as well as social sciences, and humanities. Based on the wide coverage in terms of subject areas and various disciplines, it is a suitable source for the recent study which attempts to investigate lexical bundles in multiple disciplines. Another reason that has been taken into consideration is the fact that *PLOS ONE* is ranked Q1 with an H-index of 332 according to the SC Imago journal rank (SJR) (Scimago Journal & Country Rank, 2020). Therefore, it is a high-ranking journal with a high impact factor as well.

Furthermore, *PLOS ONE* journal also has a wide range of scope, including primary research, systematic reviews, qualitative research, etc., and has been indexed by some indexing services, such as PubMed, GeoRef, Google Scholar, Scopus, Web of Science, etc. (PLOS ONE, 2022). Another important factor is the fact that *PLOS ONE* journal has ranked 9/110 in the category of multidisciplinary journals indexed by Scopus (Scopus, 2020). In addition, all articles published in *PLOS ONE* journal are written in English, which fits the criterion set by the author since only articles written in English are used in the current study.

The journal articles of the present study are taken from four different disciplines based on the classification of *PLOS ONE* journal (i.e., Chemistry, Mathematics, Psychology, and Sociology) which are commonly categorized in two different majors or wide academic disciplines, namely physical sciences and social sciences. These four disciplines are selected since they have the highest number of journal articles based on their respective subject areas. Chemistry and Mathematics, which have 43,749 and 32,736 articles respectively, are regarded as the top two disciplines with the highest number of journal articles in the subject area of physical sciences, which contains 90,262 articles in total.

Meanwhile, Psychology (i.e., 41,917 articles) and Sociology (i.e., 14,062 articles) are also the top two disciplines with the highest number of journal articles in the subject area of social sciences, which contains 58,120 articles in total. Therefore, these four disciplines are selected to represent two different subject areas in *PLOS ONE* journal: physical sciences and social sciences. By selecting these four disciplines, the present study attempts to compare the characteristics of these

multiple disciplines in terms of their lexical bundles, as well as to find out the distinctiveness of two different major disciplines (i.e., physical sciences and social sciences).

Many previous studies have shown that lexical bundles vary among disciplines (e.g., Kwary et al., 2017; Ren, 2021; Yin & Li, 2021). Some of these studies, for instance, Ren (2021) and Yin & Li (2021) used research articles from different academic journals based on the selected respective disciplines. The variation of lexical bundles in these studies is assumed to be affected by academic journals that have been chosen since every academic journal has its characteristics. Therefore, investigating how lexical bundles vary among the multiple disciplines based on journal articles taken from one certain multidisciplinary or even interdisciplinary journal may bring new insights into the study of lexical bundles. Based on this assumption, the present study needs to be conducted.

1.2 Research Questions

Based on the background of the study that has been stated, this thesis, therefore, attempts to fill in the gap by raising the following questions:

1. How can lexical bundles become a marker to distinguish the distinctiveness of journal articles in the four disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) taken from one multidisciplinary or interdisciplinary journal based on the structural forms and functional classifications?
2. How does the semantic preference, as the extended analysis of functional classifications, determine the characteristics of journal articles in the four

disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) based on their shared bundle(s)?

1.3 Objectives of the Study

The current study has two objectives which are reflected by the research questions, such as:

1. To determine how lexical bundles can become a marker to distinguish the distinctiveness of journal articles in the four disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) by investigating their structural forms and functional classifications;
2. To find out how the semantic preference determines the characteristics of journal articles in four disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) based on their shared bundle(s).

1.4 Significance of the Study

The present study is theoretically expected to contribute to the field of vocabulary learning and phraseology by providing the list of lexical bundles which frequently occur in journal articles related to the four disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology). Phraseology, according to Collins dictionary, is defined as follows: (1) the manner in which words or phrases are used, and (2) a set of phrases used by a particular group of people (Collins, 2022). Besides, by investigating the lexical bundles based on their structural forms and functional classifications, the study is expected to become an essential component in the success of language learning among academicians/researchers, especially for the students and lecturers who are regarded as novice learners and non-native

speakers that are willing to improve and enhance their knowledge about vocabulary and phraseology.

Furthermore, these findings will also help them to understand the characteristics of the four disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) based on the different forms and use of lexical bundles. In addition, by extending the analysis of functional classifications through the analysis of semantic preference, this study attempts to determine more specific characteristics of journal articles based on their shared bundle(s). All these efforts are practically expected to help and assist the academicians/researchers, especially students and lecturers of these related disciplines, in writing and publishing articles in reputable international journals.

CHAPTER II

LITERATURE REVIEW

This chapter presents the theoretical framework which consists of some linguistics theories that have been used to analyze the lexical bundles in multiple disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology). In addition, the chapter also describes some lexical bundle studies in journal articles, particularly research articles of several academic disciplines.

2.1 Theoretical Framework

This sub-chapter consists of several theories that have been applied to support the analysis of the study, such as corpus linguistics, lexical bundles, and semantic preference. All theories that have been explained in this chapter are combined with ideas to produce comprehensive research in the field of linguistics, particularly in the study of lexical bundles in journal articles of multiple disciplines.

2.1.1 Corpus Linguistics

Before jumping to the definition of corpus linguistics, it is important to define what the term corpus (pl. corpora) is. According to Cheng (2012), a corpus is a collection of texts that has been gathered based on particular design criteria for a specific reason. Therefore, corpus linguistics is the study of language which involves a wide range of data in the form of a massive compilation of written text and the assistance of computer analysis (McEnery & Hardie, 2012).

Despite the definition, the perspective, or the way how scholars comprehend corpus linguistics are different. Some linguists consider corpus as a theory (Tognini-Bonelli, 2001; Cheng, 2012) while, others regard it as a methodology,

meaning that it is not the same as linguistics branches, such as sociolinguistics, psycholinguistics, semantics, pragmatics, etc. As one of many scholars who stands with the former notion, Tognini-Bonelli (2001) has argued that corpus linguistics is a pre-application methodology, which means that it can determine sets of rules and pieces of knowledge on its own before they are applied. Therefore, it is clear that corpus linguistics has a theoretical status and can contribute particularly to other applications.

Some scholars who stand with the latter notion state that corpus does not explain what is studied, and therefore, it is a methodology that comprises numerous related methods that can be employed by scholars of many different theoretical learnings (Lindquist, 2009). Furthermore, It has been argued that corpus linguistics is a methodology that enables us to improve insights into how language works by 'consulting' data based on a real-life situation. Thus, it is moderately clear that corpus research cannot be carried out on a theoretical basis solely (Weisser, 2016).

Even though the debate on this notion may never be fully ended, these different standpoints, as stated by Cheng (2012), should be highlighted since whether the position a person takes will more or less affect the approach that they are going to apply in linguistics study. As has been asserted by Cheng (2012), those who see corpus linguistics as a methodology use the term 'corpus-based approach'. In this approach, corpus linguistics is used to examine present theories or frameworks against evidence in the corpus. On the other hand, corpus linguists who regard it as a theory use the term 'corpus-driven approach', which considers the corpus as the starting point to develop theories about language.

The distinction of corpus linguistics whether it is a theory or a method is proposed by Tognini-Bonelli (2001). In the corpus-driven approach, Tognini-Bonelli (2001) mentions that the corpus is regarded as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system. Besides, the linguists in this approach are committed to the integrity of the entire data, and the descriptions are devoted comprehensively according to the evidence of the corpus.

On the contrary, linguists who stand with a corpus as a method are familiar with the corpus-based approach, which according to McEnery & Hardie (2012), rely on the corpus data to investigate a theory or hypothesis. The corpus data in this approach are used to validate, refute, or refine the existing/established theory or hypothesis in the current literature. In other words, it can be asserted that while the corpus-as-method regards corpora and corpus techniques as sources of empirical data that can be employed to support or reject any theory that explains language, corpus-as-theory refuses any language patterns explanation which does not obtain directly from the interaction between the analysts with their data (McEnery & Hardie, 2012).

As for the present study, the corpus data are used to evaluate and validate results that have been shown by some lexical bundle studies. Furthermore, the present study also extends the analysis of discourse functions by applying semantic preference, which is determined based on the similarity of semantic features. Therefore, in this case, this study applies a corpus-based approach. At the same

time, the present study can also be regarded as a corpus-driven study since it employs a bottom-up research model.

2.1.1.1 Concordance

Concordance is a feature that is commonly used in the process of text analysis in a corpus. It is defined as the collection of all search items or all contexts that have been displayed in a corpus software tool or computer program (e.g., AntConc) in the form of KWIC (i.e., Key Word in Context), which is located in the center along with its co-text that is located to the left and right (Cheng, 2012; Lindquist, 2019).

In the process of investigating the lexical bundles, conducting the concordance analysis is a very fundamental process since the functional classifications of each bundle are determined based on the information provided by the concordance. The KWIC of the corpus in this study is all types of lexical bundles which have occurred and sorted in all multiple disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology). The functional classifications of these bundles are determined by looking at the co-texts which are on the left and the right of each bundle. In addition, the concordance analysis is also used to analyze the semantic preferences of the current study since it enables the author to pay attention to the preceding and the following collocations which complement the collocate words, or in this context, are the lexical bundles.

2.1.2 Lexical Bundles

The majority of studies devoted to examining lexical bundles have proved that lexical bundles differ across registers (e.g., Biber et al., 2004; Biber, 2006);

genre (e.g. Hyland, 2008a, 2008b); disciplines (e.g., Hyland, 2008b, 2012; Hyland & Jiang, 2018) and writers, such as L1 and L2 English (Pérez-Llantada, 2014; Bychkovska & Lee, 2017; Appel & Murray, 2020). Based on these studies, it has been found that there are several characteristics of lexical bundles that can be highlighted. First, even though lexical bundles are commonly known as neither complete structural units nor idiomatic, they are regarded as the building blocks of discourse (Biber et al., 1999; Biber et al., 2004).

Second, lexical bundles are regarded as a strong element to distinguish discipline, which has been proved by their variation in terms of structural forms and functional classifications (Hyland, 2008b; Kwary et al., 2017; Ren, 2021; Yin & Li, 2021). Third, the use of lexical bundles adequately and appropriately is not only an indicator to determine language proficiency (Wei & Lei, 2011), but also a marker to determine expertise in writing, as well as membership in the discourse community (Salazar, 2014). Numerous studies focused on examining lexical bundles have also made it clear that three criteria should be fulfilled in considering multi-word expressions as lexical bundles, such as length or sequence; cut-off frequency and minimum frequency or frequency threshold; and range or document frequency.

The sequence of bundles can be in the form of at least two-word, three-word, four-word, etc. Two-word bundles are claimed to be extremely common and not useful in the research of diachronic corpus (Kopaczyk, 2012 as cited in Hyland & Jiang, 2018). On the other hand, five-word and six-word bundles are considered to be rarer than other sequences and often contain shorter bundles (i.e., two-word,

three-word, and four-word bundles) (Hyland & Jiang, 2018). Four-word bundles are sequences or strings of bundles that are regarded as the most often studied bundles in academic discourse (Hyland & Jiang, 2018).

It is because four-word bundles are ten times more frequent and provide a wider variety in terms of structures and functions to examine compared to five-word bundles (Biber et al., 1999). Furthermore, even though three-word bundles are ten times more common than four-word bundles (Biber et al., 1999), the three-word bundles are subsumed in the structures of the four-word bundles (Cortes, 2004). In addition, Hyland (2008b) also mentions that the four-word bundles provide a clearer range in terms of structures and functions compared to three-word bundles.

Similar to the sequences or strings of words, the minimum frequency or frequency threshold and the range or document frequency of lexical bundles also vary. The frequency threshold is commonly set to at least 10 times of occurrences per million words (Biber et al., 1999; Biber, 2006); 20 times per million words (Cortes, 2004; Hyland, 2008a, 2008b); 40 times per millions of words (Biber et al., 2004); and even raw frequencies (Chen & Baker, 2010). On the other hand, the range of lexical bundles is set to either must be occurred in at least 5 different texts (Biber et al., 1999) or 10% of texts (Hyland, 2008a, 2008b). This criterion must be fulfilled to avoid individual speakers' or writers' idiosyncrasies or quirks (Biber et al., 1999; Hyland & Jiang, 2018). The analysis of lexical bundles mostly concerns the structural forms and functional classifications which will be explained more in the sub-chapters below.

2.1.2.1 Structural Forms of Lexical Bundles

The current study applies the structural analysis of lexical bundles based on the taxonomy proposed by Biber et al. (1999) since it is a comprehensive classification, a reliable all-rounded frame, as well as a starting point to classify the structural forms of bundles (Yin & Li, 2021). Furthermore, the taxonomy has also been applied in numerous studies (e.g., Kwary et al., 2017; Damchevska, 2019; Yin & Li, 2021, etc.). Biber et al. (1999) categorize lexical bundles in academic prose based on their structure correlates into 12 sub-classes, which can be distinguished into 3 broader categories, namely bundles in the form of the noun phrase (NP), such as “*the results of the*”, “*the fact that the*”, etc.; prepositional phrase (PP), such as “*as a result of*”, “*in the present study*”, etc.; and verb phrase (VP), such as “*it is important to*”, “*is one of the*”, etc. (Chen & Baker, 2010). In addition, there are also other expressions, which are defined as lexical bundles that cannot be classified, and therefore, do not belong to the previous categories, such as “*as well as the*”.

2.1.2.2 Functional Classifications of Lexical Bundles

Biber (2006) argues that discourse functions are an essential element to obtain a better understanding of the distribution of lexical bundles. Therefore, the taxonomy of functional classifications has been proposed to classify these multi-word expressions properly. Biber (2006) distinguishes the functions of lexical bundles into three major classifications, such as stance expressions, discourse organizers, and referential expressions. Stance expressions are used to communicate certainty or possibility assessment, as well as to express attitudes in

certain propositions, such as “*the fact that the*”, “*can be used to*”, “*it is possible that*”, etc. Meanwhile, discourse organizers are used to indicating the correlation between the preceding and following discourse, such as “*in the present study*”, “*on the one hand*”, “*on the other hand*”, etc. (Biber, 2006; Salazar, 2014). Additionally, referential bundles are used to point at certain references, such as time, place, text, etc.; concrete or abstract things; as well as quantity, for instance “*is one of the*”, “*the size of the*”, “*the extent to which*”, “*as can be seen*”, etc.

2.1.3. Semantic Preference

Shin (2020) mentions that the analysis of semantic prosody and semantic preference in the argumentative essays written by L1 and L2 students can be used to compare writing produced by different language groups. Therefore, following the study conducted by Shin (2018, 2020), the current study also extends the functional classifications analysis of lexical bundles by investigating the semantic domains of lexical bundles in journal articles of multiple disciplines. It is based on the assumption that if the functional classifications, which have been commonly carried out in the study of lexical bundles, show that the use of lexical bundles varies in different disciplines, then it is obvious to assume that focusing the investigation on the semantic domains, as an effort to extend the functional classifications analysis, may also become other remarkable elements to determine the use of lexical bundles in journal articles among different academic disciplines, as well as to bring another new perspective in the variation of lexical bundles analysis.

However, unlike Shin (2018, 2020) who investigates both semantic prosody and semantic preference in the study of lexical bundles, the present study only

focuses on examining the semantic preference of the shared bundle(s) among the corpora of the four disciplines (Chemistry, Mathematics, Psychology, and Sociology). The main reason is that the identification of semantic prosody of words expressions, particularly lexical bundles in the current study, will be based on the way how the analyst determines them either positive, negative, or even neutral, which is quite difficult to be carried out objectively (Bednarek, 2008). On the other hand, Bednarek (2008) also mentions that even though the term semantic prosody is usually associated with semantic preference, the semantic preference of word expressions or lexical bundles is relatively observable and can be conducted objectively based on the corpus evidence.

Besides the subjectivity of the semantic prosody itself, another reason that has been taken into account is the fact that the corpora of the current study consist of journal articles in multiple disciplines. It is not sufficient for the author to interpret the prosody of lexical bundles based on his knowledge solely, since the author does not have any background knowledge regarding the multiple disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) that have been selected. Therefore, the semantic prosody must be determined with the help of experts who have background knowledge of these respective disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) to understand the journal articles comprehensively and also to obtain excellent results of interpretation.

Semantic preference is defined as a connection between a lemma or word form and a set of words that are semantically related (Stubbs, 2001). In other words, semantic preference categorizes elements based on semantic similarity (McEnery

& Hardie, 2012). Besides, according to McEnery & Hardie (2012), the similarity between semantic preference and semantic prosody can be found in the fact that these two are abstractions across multiple, different collocations. One factor that can distinguish these two terms is if a semantic preference may support any field of definable semantics, semantic prosody, on the contrary, is devoted to evaluation, either positive or negative.

In addition, semantic preference connects the node to some words in its context based on a specific field of semantics (i.e., word meaning). Meanwhile, semantic prosody connects the node to several attitude expressions or evaluations which may not be in the form of a single word but may be given in the larger context, such as in the discourse and pragmatics field as stated by McEnery & Hardie (2012:138). The interpretation of semantic prosody, which is certainly beyond the semantic field is also in line with what has been stated by Stubbs (2001:65-66). However, instead of using the term semantic prosody, Stubbs proposed the term *discourse prosody*, which has been mentioned to be concerned with the attitude of the speaker or speaker meaning instead of the word meaning.

It is clear that when the interpretation involves the attitude or expressions of the speaker, the analysis is certainly at the level of the pragmatic domain rather than the semantic domain (McEnery & Hardie, 2012). Since the interpretation of the data in the present study is conducted based on the corpus evidence solely, the extended analysis of lexical bundles will be particularly restricted to semantic preference which concerns the semantic fields or word meaning (McEnery & Hardie, 2012).

One of the examples of semantic preference is investigated by Stubbs (2001), who finds that *undergo* is commonly related to categories of medical (e.g., surgery, treatment, medical, operation, and heart); training and testing (e.g., training, examination and test(s)); and change (e.g., change(s) and transformations). Another example is the investigation conducted by Partington (2004) who finds that *completely*, *entirely*, *totally*, and *utterly* are associated with the "absence of a quality" or "change of state". Collocates which fall into the former category are *helpless*, *useless*, *unable*, and *forgotten*. Meanwhile, collocates that fall into the latter category are *changed* and *different*.

2.2 Related Studies

In this subchapter, the author reviews some studies that are closely related to the current study. The related studies described here are mostly articles that are concerned with investigating lexical bundles in research articles of one certain discipline and multiple disciplines either in the whole part of the research articles or in a particular part of the articles (e.g., abstracts). All studies included here are written and published in the last decade. Additionally, there is also one study that focuses on argumentative essays.

The first study was the investigation of lexical bundles in the research articles of Economics conducted by Damchevska (2019). Focusing on the four-word bundles, the study attempted to analyze the structure of lexical bundles in the research articles on Economics written by Macedonian scholars, that had been published between 2011 and 2017 in two different Economics international journals based in Skopje, Macedonia: *Economic Development and CEO journal of*

Economics. There were 60 articles taken from the Economic Development journal which comprise 188,903-word tokens and 8,128-word types. Meanwhile, there were 40 articles taken from the CEA Journal of Economics which comprise 189,299-word tokens and 8,356-word types. Therefore, the corpus consisted of 100 articles which comprise 378,202 tokens.

The cut-off frequency was set to 20 times per million words and the frequency threshold was based on raw frequency, i.e., 7.56, which was rounded up to 8. On the other hand, the dispersion or range was set to 10 articles or 10% of the texts. The study revealed that there were 141 types of bundles occurred in the study. These findings were then compared with the study conducted by Hyland (2008b). Based on the comparison, it was found that there were 40 shared bundles between Damchevska's corpus of Economics (Economics research articles corpus, henceforth ERAC) and Hyland's overall corpus. Besides, there were 49 shared bundles found between ERAC (Damchevska, 2019) and the social sciences corpora (Hyland, 2008b), and 31 shared bundles found between ERAC and pure and applied sciences corpora (Hyland, 2008b).

As for the structural form, it was found that based on the taxonomy of Biber et al. (1999), the types of bundles in ERAC were dominated by NP + of, other PP, and PP + of with a percentage of 29.08%, 17.74%, and 15.61% respectively. In contrast, anticipatory it, passive verb + PP fragment, copula *be* + noun phrase, Verb/adjective + to-clause fragment, and Pronoun/noun phrase + *be* + *were* regarded as the lowest structural forms in this study with the total percentage of 4.25%, 2.84% and 0.70%, respectively.

What makes the current study differs from this related study is the number of disciplines that are involved. Instead of focusing on one discipline, the current study compares four different disciplines (i.e., Chemistry, Mathematics, Psychology, and Sociology) which can also be regarded as the comparison of two different major or wide disciplines. It is because Chemistry and Mathematics can be categorized as physical sciences, while Psychology and Sociology are classified as social sciences. In addition, putting the comparison of ERAC and what had been conducted by Hyland (2008b) aside, the focus of this study is the structural forms of lexical bundles (Biber et al., 1999) which is only the surface part of the lexical bundles' study. On the contrary, the current study also investigates the functional classifications of lexical bundles based on the taxonomy proposed by Biber (2006).

The next study was conducted by Kwary et al. (2017), which examined lexical bundles in journal articles of multiple major or wide disciplines. This study concerned with the structural forms and functional classifications by employing a corpus of four different wide academic disciplines, namely health sciences, life sciences, physical sciences, and social sciences based on the classification of Scopus (Elsevier, 2016 as cited in Kwary et al., 2017). Besides, the corpora comprised 2,937,431 words and were collected from 400 articles. There were some criteria that the authors set in selecting the articles to be used in creating the corpus. First, the articles must be open access articles that could be accessed and downloaded easily. Second, the articles used as the data were written in English. Third, the article must be written by at least one English native speaker, which could

be confirmed based on the name of the authors. Finally, the selected journals must have the highest 5 years impact factors.

The criteria of lexical bundles were four-word bundles which according to Hyland (2008b) are more common than the five-word bundles, and provide a clearer range of structures and functions compared to three-word bundles. Besides, the minimum frequency was set to 20 times of occurrences, as well as the minimum range, meaning that the sequences should occur in 20 different texts. The structural forms and functional classifications were based on the taxonomy proposed by Biber et al. (1999) and Biber (2006), respectively. In addition, the lexical bundles were extracted using AntConc 3.4.4 (Anthony, 2014 as cited in Kwary et al., 2017).

The study revealed that physical sciences featured the greatest number of lexical bundles (i.e., 43). On the other hand, health sciences had the least number of lexical bundles (i.e., 3). The occurrences of lexical bundles in social sciences and life sciences were 27 and 12, respectively. Besides, the results of pair-up major disciplines also varied. The study found that the pair-up of physical sciences and life science had the greatest number of shared bundles (i.e., 8), followed by life sciences and social sciences (i.e., 6), and health sciences and life sciences (i.e., 1). In contrast, there were no shared bundles in the pair-up disciplines of health sciences and physical sciences, and neither between health sciences and social sciences.

Furthermore, prepositional-based (i.e., 37.10%) and verb-based bundles (i.e., 37.10%) dominated the distribution of lexical bundles in terms of structures among the cross-disciplines. On the contrary, the noun-based was the lowest

number of structural forms with only approximately 25.81%. In addition, in terms of functional classifications, there were 40 lexical bundles occurred in the referential expressions. The rest major categories were the discourse organizers followed by the stance expressions with 12 and 10 occurrences in total, respectively.

The domination of referential expressions was related to the needs of academic articles in identifying entities, e.g., to refer to theory, concepts, data, and findings of the research. Meanwhile, the low frequency of the stance expressions indicated that authors of academic articles tended not to express their attitudes and assessments too often in their writings.

The strength of this article lies in the fact that it uses multiple wide or major academic disciplines which certainly cover several specific disciplines, for instance, health science consists of some specific disciplines such as Dentistry, Medicine, Nursing, etc. (Elsevier, 2022). It is because numerous studies of lexical bundles in journal articles, particularly research articles, are commonly concerned with more specific disciplines, for instance, the comparison of lexical bundles in Finance, Accounting, Biology, and Applied Linguistics (Yin & Li, 2021). Unfortunately, the study did not really explain how many specific disciplines were actually included. In other words, the authors did not clearly mention how many specific disciplines, as parts of the wide or major academic disciplines, were chosen to build the corpus.

Since the corpus is based on wide academic disciplines, it is obvious to assume that the articles are collected from more than one specific discipline. However, the question of whether the authors collected the articles from all specific