A Corpus-driven Study on the Use of Passive Verb Bundles in Academic Writing: A Comparison between L1 and L2 English Speakers

Radika Subramaniam [a]
radika5352@gmail.com
University of Malaya, Malaysia

Sheena Kaur [b]
sheena@um.edu.my
University of Malaya, Malaysia

ABSTRACT

Lexical bundles studies are still limited in relation to the use of verb phrase-based bundles, specifically the passive forms in academic writing. This study reveals the lexico-grammatical patterns and functional categories of passive verb bundles in the 409,373-word Malaysian Polytechnics Electronic Engineering Learner Corpus (MyPolyEELC) in comparison to the British Academic Written English (BAWE) 29,530-word sub-corpus. The Sketch Engine lexical computing tool is used to analyse both the L2 Malaysian and L1 English learner writing. Biber et al.'s (1999) structural categories framework is employed to identify the verb-phrase bundles and further locate the passive forms, whereas the functional categories are identified through Hyland's (2008a) framework. Francis et al.'s (1996) verb pattern analysis is used to identify the lexico-grammatical patterns and meaning groups of the passive verb bundles. The analyses indicate several passive verb structural categories in both learner corpora. The L2 learners disclose limited participant-oriented functions in the passive verb bundles' use, which is contrary to the L1 English learners who display profound use of participant-oriented 'engagement' and 'stance' categories through several choices of lexical verbs. The findings also show the inter-relation between the meaning group, lexico-grammatical pattern and functional category. The inter-relationship illustrates that the meaning group which motivates the lexico-grammatical pattern or vice versa, further contributes to identification of the functional category of the bundles based on a larger context of the text. The variations discovered between the L2 and L1 learner writing may offer several corpus-related pedagogical practices to be applied in an EAP classroom.

Keywords: Academic Writing; Corpus linguistics; Engineering; Lexico-grammatical pattern; Passive verb bundles

INTRODUCTION

Lexical bundles or multi-word units (n-grams) are sequence of words which become the "basic building blocks of a discourse" (Biber et al., 1999; Biber et al., 2004, p. 371) and often addressed as 'clusters' (Scott, 1997; Hyland, 2008a). Non-native speakers of English often find lexical bundles challenging and the knowledge which one has in utilising these recurring phraseological units is recognised as a native-like proficiency (Biber & Barbieri, 2007).

Due to the significance of lexical bundles in written academic registers and discourse, a large amount of research in English for Academic Purposes (EAP) has been devoted to their use (see Chen & Baker, 2010; Durrant, 2017; Hussain et al., 2021; McEnery et al., 2019). Among all, the largest database has been used by Biber et al. (1999, p. 997) in researching the

[a] *Main author*
[b] *Corresponding author*

eISSN: 2550-2131
ISSN: 1675-8021

spoken and written academic registers based on the 40-million-word Longman Spoken and Written English (LSWE). The study illustrated twelve structural categories which occur in written academic registers. These structural categories are associated with noun phrase, verb phrase, prepositional phrase and other expressions. This shows that not only lexical bundles are important, but also exist in almost all phrasal structures.

Various approaches have been taken to study lexical patterns in learner corpora. Studies have compared the use of lexical features between native and non-speaker learner corpora in Master's Dissertation, PhD Thesis, English language test and essays (e.g., Adel & Erman, 2012; Chen & Baker, 2010; McEnery et al., 2019). The structural and functional categories of multi-word unit items have also been explored via comparison with experts' use of language in published research articles to see how the variation of lexical bundles occurs between novice writers and expert writers (Chen & Baker, 2010; Liardet & Black, 2019).

A review of these studies reveals that while lexical bundle features in academic writing have been extensively studied, EAP studies have pointed out the limitations of research on the use of these word sequences among L2 learners. In a study, which was conducted between the novice L1 and L2 writers in comparison to experts' writing, it was found that verb phrase-based bundles appear more frequently in novice writer's piece of writing with L2 learners' generally lower rate of passive verb bundles' use than the L1 English novice writers (Chen & Baker, 2010). This finding demands a deeper investigation into the verb phrase-based bundle patterns especially passive forms among L2 learner writers.

More relevant to the current research are studies within EAP that have also investigated the use of multi-word units (i.e., collocations, prefabrications, idioms) in academic discourse associated with engineering discipline (e.g., Hyland, 2008a; Mudraya, 2006). Within the Electrical Engineering domain, the 'passive + prepositional phrase (PP)' bundles were found to be recurring more frequently in academic writing than any other structural categories (Hyland, 2008a). Likewise, the 'passive + PP' bundles occupy at least 12% to 17% of the total percentage of structural categories which occur in Electronic Engineering sub-disciplines (i.e., Power, Telecommunication, Control and Computer) (Rezoug & Vincent, 2018). These findings should not be neglected but further investigated to realise more types of passive verb patterns in the Electronic Engineering written academic text.

Corpus studies reporting the use of lexical bundles in academic texts of engineering learners at Undergraduate, Master's and PhD level are often cited when exploring the L2 learner language (e.g., Hyland, 2008a; Rezoug & Vincent, 2018). However, lexical bundles studies are still limited in relation to the use of verb phrase-based bundles, specifically the passive forms in academic writing. What is more, studies in the use of lexical bundles among L2 learners at a Diploma level (i.e., post-secondary) in the Technical and Vocational Education Training (TVET) institutions have not been studied to date.

Likewise, in the Malaysian context, thus far, corpus-based studies have only explored ESL learners' writing in argumentative essays (e.g., Roslina Abdul Aziz, 2018; Siti Aeisha Joharry, 2016) without investigating the discipline-specific texts. As the Diploma level learners at Malaysian Polytechnics receive less formal exposure to the academic writing in English language, these learners may find it difficult to compose appropriate passive sentences when writing the project report. To fill this gap, the Technical and Vocational Education Training (TVET) learners' use of verb phrase-based bundles especially passive forms in the Final Year Project (FYP) report writing at Malaysian Polytechnics in the field of Electronic Engineering is investigated by the current study.

eISSN: 2550-2131
ISSN: 1675-8021

PASSIVE CONSTRUCTION IN ACADEMIC WRITING

Passive voice occurs more commonly in written academic prose in comparison to news and conversation (Biber & Conrad, 2019). Passive construction analysis especially in technical texts is no less important than any other grammatical structure analysis as 'process' plays a major role in technical writing and to describe the process flow, passive is apt (Swales & Feak, 2004). In the context of the present study, the writing of the FYP report necessitates the learners to explain the processes and procedures which have been implemented throughout the project completion using passive forms (Teh et al., 2016). This requirement may not be fulfilled by every student by making appropriate use of passive structures, as the TVET students are second language learners who generally possess low to average English language proficiency.

It was reported that the teaching and learning of passive structure is not an easy task especially for L2 learners as it is one of the most problematic language features especially for speakers of Asian languages (Hinkel, 2002). More recently, Granger (2013) in a comparative study between the native and non-native learner writing identifies the underuse of passive among the L2 English learners. Interestingly, the use of passive voice grows in line with the level of proficiency among the learners, which means, the passive structures are more prevalent in the academic writing of more advanced learners with higher language proficiency, and it is associated with high-scoring texts which statistically signifies strong positive correlation between the use of passive and improved writing (Biber & Gray, 2013).

According to Quirk et al. (1985, p. 159), there are two grammatical levels which need to be addressed in analysing passive voice, which are the "verb phrase" level and "clausal level. The verb phrase level requires some transformation of the main verb into its past participle form with a 'be' form of auxiliary preceding it (e.g. *is used*, *are connected*). On the other hand, the clause level structure demands more than a phrase level by involving changes to the other constituents within the sentence where the subject and object are relocated (e.g., Active: *We used Arduino to develop the project* / Passive: *Arduino was used to develop the project*). Here the "active subject becomes the passive agent" and the "active object becomes the passive subject" with a *by*-prepositional phrase introduced (optional) (Quirk et al., 1985, p. 159). In this study, finite verbs which occur in the format of a 'be' verb or any other auxiliaries including a modal verb followed by a past participle and non-finite past participle verbs are defined as passive verb categories. This parameter is set as one of the operational definitions to work on analysing the passive-related verb phrase bundles in the learner corpora. However, since there is interface between lexis and grammar, as indicated in the following section, analysis at higher levels needs to be taken into consideration.

THE INTERFACE BETWEEN LEXIS AND GRAMMAR

Lexico-grammatical analysis incorporates both lexis and grammar to see how they affect the use of language in any context. The occurrence of certain lexical items within certain grammatical structures is not merely arbitrary but it is justifiable, why certain patterns occur frequently in such associations (Biber et al., 1999; Staples & Reppen, 2016). The presence of certain lexico-grammatical features in different texts is motivated by different communicative purposes (Biber & Conrad, 2019). In this study, lexico-grammar is defined within the corpus-driven grammar perspective. The lexico-grammatical pattern in Corpus Linguistics studies is revealed through the corpus-driven grammar (i.e., pattern grammar) to see the occurrence of target lexical items or bundles syntagmatically or paradigmatically in a text (Francis, 1993). This pattern is defined as "a phraseology which is frequently associated with (a sense of) a word, in terms of the prepositions, groups and clauses that follow the word where each pattern occurs with a restricted set of lexical items and each lexical item occurs with a restricted set of

eISSN: 2550-2131
ISSN: 1675-8021

patterns" (Hunston & Francis, 2000, p. 3). The pattern grammar approach provides a pedagogy-oriented analysis of a lexical item/bundle by identifying the associating grammatical or lexical combinations as the pattern inventory can be used as a source of reference in teaching and learning (Green, 2019).

Studies have reported the importance of highlighting the lexico-grammatical patterns in written academic texts to enhance learner writing (e.g., Biber & Gray, 2013; Green, 2019; Lim, 2017; Qian & Ma, 2020; Staples & Reppen, 2016). Green (2019) introduces Lexicogrammar of Academic Vocabulary List (LAV) derived from a 9.3-million-word corpus of textbooks from various disciplines (i.e., Physics, Geography, Biology, Chemistry, Mathematics, Economics, English and History). These lexico-grammatical patterns of the content words are realised through pattern grammar approach, however without exploring the semantic properties of the vocabulary list. Likewise, Qian and Ma (2020) disclose the pattern and meaning relationship by introducing a grammar pattern list based on frequently occurring academic verbs to ease the learning of discipline-specific grammar patterns without taking the semantic properties into account.

Given that the linguistic phrase is "the place where structures are engineered to allow meaning to take shape" (Sinclair, 2009, p. 408), the study of phraseological pattern, does not merely involve lexical and grammatical aspects, but also its semantic properties which emerge from theory of contextual meaning (see Firth, 1957). In line with this perspective, Gries (2008) identifies that the phrase clusters are related both syntactically and semantically under the lexico-grammatical framework. Likewise, Sinclair (2005) points out that meaning is significant in phraseological unit analysis as each pattern carries its meaning based on the co-occurring companies and the grammatical structure, hence meaning is identified through the phrase but not the occurrence of a single word. In Sinclair's (1996, 2005) model of extended lexical unit (ELU), lexico-grammatical properties and its meaning is realised through four parameters, which are collocation, colligation, semantic preference and semantic prosody. This model discloses the meaning from concrete to abstract by exploring the co-occurring content and grammatical words, syntagmatically (horizontally in a corpus concordance output) through collocations and colligations, paradigmatically (vertically in a corpus concordance output) through semantic/topic relation and pragmatically (the whole text) through semantic prosody which illustrates the communicative purpose / function of the text (Stubbs, 2009).

The interrelation between pattern (syntactic element) and meaning (sense) has been revisited from Sinclair's lexical grammar by Francis et al. (1996), by proposing meaning groups of nouns, adjectives and verbs via Collins COBUILD Grammar Patterns Series (Hunston & Francis, 2000). Francis et al. (1996, p. ix) divided "the verbs with each structure into groups according to their basic meaning". Generally, the verb patterns are grouped based on the following (Francis et al., 1996, p. xxv-xxvi), as presented in Table 1:

TABLE 1. Verb patterns

Category	Sample verb patterns
Simple patterns	V, V n, V adj, V –ing, V to-inf
Simple patterns with prepositions and adverbs	V prep/adv, V adv/prep
Complex patterns	V n n, V n adj, V n –ed
Complex patterns with prepositions and adverbs	V n prep/adv, V n adv/prep
Patterns with 'it'	*it* V clause, V *it* clause, *it* V, *it* V adj, V *it*
Patterns with there	*there* V n, *there* V n prep/adv
Combination of patterns	V *about* n, V *at* n, V with quote

Passivization of the same pattern (i.e., the active forms) sees changes to the main verb form as it becomes the past participle (i.e., 'V-ed' (Hunston & Francis, 2000, p. 200), as shown in Table 2:

TABLE 2. Active and passive verb patterns

Active	Passive
V n *to* n	*be* V-ed *to* n
V n n	*be* V-ed n

The same pattern which appears with different lexical verbs is categorised into either the same or different meaning groups (Francis et al., 1996). The same meaning group means that the verb can be grouped together based on its sense and syntactic relationship (e.g., the lexical verb '*see*' as in the structural pattern of '*it can be seen that*' and '*note*' as in '*it should be noted*' share the same meaning group, that is 'think and discover') (Francis et al., 1996, p. 527). Studying both the syntactic and semantic patterns of the verb based on its co-occurring partners is sensible in exposing the lexico-grammatical properties the verb postulates, to further understand the communicative purpose of the text. In this study, the pattern and meaning groups are observed to postulate the lexico-grammatical patterns of the passive verb bundles. The analysis is also substantiated with the identification of the functional categories of the bundles as further elaborated in the following section.

FUNCTIONAL CATEGORIES

Hyland (2008a, p. 13-14) divides the functions of lexical bundles in dissertations and research articles in relation to three categories which are research-oriented, text-oriented and participant-oriented. Research-oriented (RO) is rather topic-oriented as it mentions the activities which involve throughout the text in relation to 'procedure', 'description', 'quantification', 'location' and 'topic', whereas text-oriented category (TO) includes bundles which help the discourse-orientation / organisation via 'transition signals', 'structuring signals', 'framing signals', and 'resultative signals'. Participant-oriented (PO) category, however, shows a reader-writer relationship by attracting the reader involvement in the text and indicating the text writer(s) involvement by expressing their opinion or judgement via two sub-categories respectively (i.e., 'engagement' and 'stance').

Research-oriented category has been identified as the most commonly occurring in academic texts in relation to Science and Technology, followed by text-oriented and participant-oriented category (Hyland, 2008a; Rezoug & Vincent, 2018). 'Resultative signals' which illustrate reporting of results or discovery have been identified more frequent in native speaker learner writing (Adel & Erman, 2012) and less frequent in L2 learner writing (Rezoug & Vincent, 2018). Hyland's (2008b) study which identifies that the use of stance and engagement features (i.e., participant-oriented category) is greater in doctoral dissertations compared to master's theses, shows the relationship between the learner level and the use of functional category. More uses of stance expressions (i.e., *it has been suggested*, *it could be argued*) have been reported in Chen and Baker (2010) by L1 English learners compared to L1 Chinese learners. Similarly, Adel and Erman (2012) have identified more types of stance bundles exploited in native speaker undergraduate linguistics students' text when comparing the corpus with L1 Swedish learners. A recent study has also found that the elements of judgement and evaluation seem to be prevalent in native speaker learner and expert writing (i.e., L1 English) with more variations of this type of stance expression (Pan & Liu, 2019). This is different from both novice and expert L2 Chinese learners who adopted less participant-

oriented approach which expresses limited writer-reader engagement in the text (Pan & Liu, 2019).

Based on these past studies, it can be observed that technical academic texts show more uses of research-oriented bundles, however the use of writer/reader-oriented expressions are somewhat limited among the non-native speaker learners and learners with a lower proficiency level as the participant-oriented feature establishes the writer's involvement in the text to present evaluative judgements. This limitation further highlights the importance of studying the use of stance and engagement features in the academic writing of L2 writers with less proficiency.

RESEARCH QUESTION

How do the lexico-grammatical patterns and functional categories of the passive verb bundles' use vary between the L2 learners in the Malaysian Polytechnics Electronic Engineering Learner Corpus (MyPolyEELC) and L1 English learners in the British Academic Written English (BAWE) sub-corpus?

METHODOLOGY

CORPORA DEVELOPMENT

The corpus for the study consists of the TVET learners' Final Year Project (FYP) reports collected from four Malaysian Polytechnics which offer Electronic Engineering programmes. As indicated by Sinclair (2005), the corpus content comprises language which is used in the target community and maximally represents the target language (Sinclair, 2005). In total, 113 project reports were collected from June 2019 and December 2019 academic sessions that later contributed to the development of a 409,373-word corpus. A complete FYP report is submitted in the format of a research report comprising 5 Chapters (Introduction, Literature Review, Methodology, Results and Discussion, and Conclusions and Recommendations) together with an 'Abstract'. The corpus is known as Malaysian Polytechnics Electronic Engineering Learner Corpus (MyPolyEELC).

As per ethical consideration, the approval from the Research and Innovation Centre of the Department of Polytechnic and Community College Education, Malaysia was obtained to collect the Electronic Engineering students' FYP reports. The Electronic Engineering programmes are offered under the Electrical Engineering Department; hence, the researcher approached the FYP Coordinators from the respective polytechnics departments to get access to the students' reports.

The corpus development included corpus cleaning where all the unnecessary features in the analysis of language samples were removed. These included diagrams, tables, appendices, equations and flowcharts with only keywords which were eliminated while transferring the content from the reports to the text files. The main headings and sub-headings were meta-tagged (indicated in the form of brackets in between text) to assist the corpus search. The Sketch Engine lexical computing tool (Kilgariff et al., 2014) was used to compile the corpus and perform the analysis to identify the multi-word unit frequency and concordance lists. The Sketch Engine comes with a built in Part of Speech Tagger (English Penn Treebank Tag set), with a list of 55 tagsets ranging from lexical categories to grammatical categories. Although the corpus is automatically Part of Speech (POS) tagged by the Sketch Engine, a manual analysis was still carried out to verify the tagging as the learner corpus may contain grammatical errors. For the BAWE corpus, the access for the corpus was obtained via the

Sketch Engine as BAWE is available as one of the preloaded corpora via the corpus analysis tool which enables its users to customise the corpus search (see Table 3).

TABLE 3. A Comparison between the MyPolyEELC and BAWE sub-corpus

Corpus / Corpus Feature	MyPolyEELC	BAWE sub-corpus
Tokens	487, 076	274, 598
Words	409, 373	229, 530
Discipline	Electronic Engineering -Communication -Computer -Control	Physical Science -Engineering -Physics -Mathematics -Computer Science -Cybernetics and Electronic Engineering
Text Genre	Final Year Project Reports	Research Report Proposal Methodology Recount Literature Survey Design Specification Case Study
Number of Texts	113 -Communication (57) -Computer (48) -Control (8)	62
Learner Language Background	Malaysian learners (L2 speakers of English)	L1 speakers of English

DATA ANALYSIS PROCEDURE

This study employs a mixed-method research design that uses a corpus-driven analytical approach (Tognini-Bonelli, 2001). Both quantitative (i.e., frequency count and log-likelihood test of significance) and qualitative methods (i.e., text analysis to identify linguistic feature) were utilised for data analysis purposes. The N-gram feature in the corpus tool was utilised for identifying the most frequently occurring lexical bundles within the MyPolyEELC and BAWE sub-corpora. Since both corpora are of different sizes, instead of the normalised frequency threshold, this study uses 6 times frequency threshold in both the corpora (regardless of size), to establish an expression as a lexical bundle. This is lower than that of Biber et al.'s (1999) work which sets 10 times per million words as the frequency threshold. A much lower threshold is used to identify the expressions which occur within both the corpora as the study focuses only on passive verb bundles as the target items for analysis and hence a minimum number of threshold may reduce the chance of overlooking passive bundles which may yield important findings in the study (Lee et al., 2020).

The other operational criteria set for analysis is the measure of dispersion between the number of texts in which the bundles occur across the corpora to avoid any skewing tendency by some writers (Pan et al., 2016). Following Biber and Barbieri (2007), the cut-off range is set at 5% from the total number of texts (i.e., 113 reports in MyPolyEELC and 62 texts in BAWE sub-corpus) which means that the passive bundles which occur at least in 5.6 (rounded to 6) and 3.1 (rounded to 3) texts respectively are considered as recurring patterns. The structural category of verb-phrase bundles is identified using Biber et al. (1999). The passive verb bundles are then short listed from the list of verb-phrase bundles identified to answer the

research questions. The structural categories of the passive verb phrase bundles discovered in the analysis are shown in Table 4:

TABLE 4. The passive structural patterns in the MyPolyEELC and BAWE sub-corpus

No.	Passive structural pattern
1	Passive verb + PP fragment
2	Verb phrase + *to*-clause
3	Verb phrase + -*ing*-adverbial clause
4	Anticipatory *it* + verb phrase
5	Verb phrase + *that*-clause
6	Adverbial clause fragment
7	*that*-relative clause

Following the pattern grammar approach (Hunston & Francis, 2000) inspired by Sinclair's (1996, 2005) lexical grammar, the lexico-grammatical patterns of the passive verb bundles are postulated using Francis et al.'s (1996) verb pattern analysis. Similar to Durrant (2017), this study employs Hyland's (2008a) framework to classify the functional categories (i.e., research-oriented, text-oriented and participant-oriented) as it is more relevant to analyse the research-based genre. The lexico-grammatical pattern and meaning group associations assist the analysis of functional categories to establish the discourse function of the passive bundles. In order to verify whether or not the difference between the bundles which appear in both the corpora are significant to be addressed for pedagogical implications, the test of significance is performed using the log-likelihood calculator available at http://ucrel.lancs.ac.uk/llwizard.html (Rayson, 2006).

IDENTIFICATION OF TARGET BUNDLES

Bundle length, frequency threshold and text dispersion are three important criteria for target bundles selection (Biber et al., 1999; Chen & Baker, 2010). Four-word length bundles are considered one of the most researched multi-word units (Chen & Baker, 2010: Durrant, 2017; Hyland, 2008a). Biber et al. (1999) on the other hand consider 3-word bundles as partly derived from longer bundles; hence bundles with longer string of words could possibly lead to more significant searches. In the present study, the search for multi-word units is set between 4 to 5-word bundles as initial analysis identified that most of the 3-word bundles occur repetitively as part of a longer string (4 and 5-word bundles). The Nest n-grams feature in Sketch Engine allows the grouping of sub n-grams which emerge from another longer string. Other than manual screening, this sub n-grams category helps the researcher to exclude repeated multi-word units as the item of analysis. As indicated by Chen and Baker (2010, p. 33), the repeated string of bundles that "overlap" could affect the results of analysis. Lexical bundles in the MyPolyEELC which occur over 6 times were identified, and the list was then scrutinised to locate all the verb phrase-based bundles using Biber et al.'s (1999) structural categories framework. These verb-based phrases were further shortlisted to identify the passive verb patterns. One longer string of words (i.e., 5-word bundle) was identified if the bundle was made up of two 4-word bundles. For example, instead of including both 4-word bundles "described in this chapter" and "is described in this", a single 5-word bundle was opted as in "is described in this chapter". There were also other instances where a single 4-word bundle was opted instead of a few 5-word bundles with similar structure, for example, "can be seen in figure", "can be seen in appendix", "can be seen in table" were categorised as "can be seen in".

Likewise, the items of analysis in the BAWE corpus were determined using the same method and criteria as in MyPolyEELC, (i.e., 4 to 5-word passive bundles which occur at least 6 times within the corpus and at least 5% of the total number of texts). However, since the

BAWE comprises a variety of texts extracted from various text genres, different learner groups (i.e., different L1 backgrounds and level of studies) and disciplines, a more personalised search was performed by creating a sub-corpus within the BAWE by selecting only relevant items to be included in the search. The BAWE consists of assignments written by Undergraduate and Master's level students from various disciplines. To ensure comparability, only undergraduates from Year 1, 2 and 3 L1 English native speaker students' work from the relevant Physical Sciences areas were selected.

Determining the target bundles for analysis is important to achieve the aim of a study, thus it includes several steps of filtering (Otto, 2021). In this study, the whole process of screening the n-grams list in order to identify the passive verb bundles, involves a few steps as follows:

1. Raw list of 4 to 5-word n-grams is generated in the Excel format**.**
2. All the bundles which meet the minimum frequency (i.e., 5 times) and text dispersion criteria (i.e., 2%) are highlighted.
3. The new list based on the criteria above is then scrutinised and all the potential passive verb bundles are highlighted.
4. These bundles are screened again to identify the final passive bundles to ensure that no repeated bundles are selected, by checking the language samples from the Sketch Engine.
5. The finalised passive verb bundles are then colour-coded based on its core lexical verb to categorise them for analysis purposes.

RESULTS AND DISCUSSION

This section provides the analysis and discussion of the findings based on similar (shared) and different (non-shared) bundles identified between the L2 and L1 learner corpora. The lexico-grammatical patterns and functional categories are presented in two sub-sections in the form of bundle groups based on the core lexical verb and function. The number indicators (e.g., [1], [2]) are provided for each pattern / bundle examples within the text to ease cross referencing of the language extracts provided in the tables (Table 5 – 8).

LEXICO-GRAMMATICAL PATTERN

'USE' BUNDLES

The 'use' passive bundles have been identified occurring in both the learner corpora. In the Student Engineering Wordlist, the past tense or past participle 'used' is the most commonly occurring form (Mudraya, 2006, p. 249-253). 'Use' in this study too, takes the significant appearance as the most frequently occurring verb in MyPolyEELC. Likewise, the BAWE sub-corpus too displays 'use' as the most common lexical verb. Despite the commonness, there is a difference in the lexico-grammatical pattern which occupies these bundles. In contrary to the BAWE sub-corpus, the prepositional fragments attached to the passive verbs in MyPolyEELC show a variety of forms, which include, 'to', 'for', 'in' and 'as' (i.e., *can be used to;* [2],[3]&[4]*can be used for*; [5]*can be used in*; [6]*can be used as*) which results in the postulation of these lexico-grammatical patterns (i.e., MODAL *be* V-ed to-inf*;* [2],[3]&[4]MODAL *be* V-ed *for* n/-ing; [5]MODAL *be* V-ed *in*; [6]MODAL *be* V-ed *as*). Variation in meaning groups which is associated with the patterns can be found, for instance [6]*can be used as* takes the 'use' meaning group that shows "the role assigned to something in the course of a particular action" (Francis et al., 1996, p. 352). On the other hand,

[2],[3]&[4]*can be used for* is connected to three meaning groups (i.e., [2]'choose and use', [3]'use' and [4]'eat') which means "appointing or choosing someone to do something or with allocating or assigning something to a particular use" (Francis et al., 1996, p. 295), "concern with using something" in general (Francis et al., 1996, p. 28) and "consuming something or not consuming something" (Francis et al., 1996, p. 20) respectively. [5]*Can be used in*, however conveys the 'use' meaning group which is concerned with using something in general. Other similar patterns associated with *used in* which can be identified in the MyPolyEELC corpus are *used in this project* [i.e., '*(be)* V-ed *in* n (non-finite) / *be* V-ed *in* n'] and *was used in this* (i.e., *be* V-ed *in* n).

The BAWE seems to be highlighting the lexical verb 'use' with 'to-inf' prepositional fragments (i.e., [1]MODAL *be* V-ed to-inf), as in [1]*can be used to* and *will be used to*. This lexico-grammatical pattern is one of the most common patterns which is utilised in both the L1 writers' and L2 writers' corpora in instances where *used to* is often utilised when specific process/processes or entity is identified to perform a particular task. In fact, *can be used to* and *used to* are the most frequently occurring 'use' bundles discovered in previous studies in relation to engineering context (Hyland, 2008a; Rezoug & Vincent, 2018).

Another form of 'use' which is explicated in both the MyPolyEELC and BAWE sub-corpus is [7] – [11]*that can be used*. In both the corpora, this bundle appears as [7]&[10]'*that* MODAL *be* V-ed to-inf' and [9]&[11]'*that* MODAL *be* V-ed prep'. However, both corpora display additional individual patterns as in [8]'*that* MODAL *be* V-ed be n' (MyPolyEELC) and [11]'*that* MODAL *be* V-ed' (BAWE). Both corpora demonstrate [7]&[10]'choose and use' as well as [9]&[11]'use' in general as the meaning groups which are equivalent to the patterns. Table 5 shows further instances of the meaning group associated with the lexico-grammatical patterns in both corpora for 'use' passive bundles.

TABLE 5. Meaning group associated with the lexico-grammatical patterns in both corpora for 'use' passive bundles.

Meaning group	**Passive verb bundles**	**Lexico-grammatical pattern**	
		MyPolyEELC	**BAWE sub-corpus**
'use' -concerned of using something in general	[5] can be used in	[5] MODAL *be* V-ed *in* n (e.g., The servo motor ***can be used in*** many ways to help robots run, to transfer remote control boats, or cars).	-
	[3] can be used for	[3] MODAL *be* V-ed *for* n (e.g., Had made some research from the journal and some solid evidence that ***can be used for*** this project).	-
	[9] & [11] that can be used	[9] *that* MODAL *be* V-ed prep (e.g., Postbox have their own features such as equipped with lock mechanism system ***that can be used*** by the house owner to keep their incoming letter).	[11] *that* MODAL *be* V-ed (e.g., The DVB standard defines the modulation and coding schemes ***that can be used***, and offers a lot of flexibility).

'use' -the role assigned to something in the course of a particular action	[6] can be used as	[6] MODAL *be* V-ed *as* n (e.g., Batteries ***can be used as*** the main power but most of smoke detectors are need to change battery every month or year).	-
'choose and use' - appointing or choosing someone to do something or with allocating or assigning something to a particular use	[1] can be used to	[1] MODAL *be* V-ed to-inf (e.g., The basic level sensor ***can be used to*** identify the point where the liquid falls below the minimum or increases above the maximum level).	[1] MODAL *be* V-ed to-inf (e.g., The process of reverse engineering ***can be used to*** replicate existing components).
	[7], [8] & [10] that can be used	[7] *that* MODAL *be* V-ed to-inf (e.g., For advanced applications, the VL53L1X supports configurable thresholds ***that can be used*** to trigger interrupts when a target is detected below a certain distance, beyond a certain distance, outside of a range, or within a range). [8] *that* MODAL *be* V-ed be n (e.g., There are many options available, but the right material ***that can be used*** is; plywood).	[10] *that* MODAL *be* V-ed to-inf (e.g., There is also a design tool designated ESDUpac A9016 which is a computer program ***that can be used*** to evaluate flat plate natural frequencies with or without in-plane loading).
	[2] can be used for	[2] MODAL *be* V-ed *for* -ing (e.g., Using proper coordination, a single transducer ***can be used for*** both emitting the pulse and receiving the echo).	-
'eat' - consuming something or not consuming something	[4] can be used for	[4] MODAL *be* V-ed *for* n (e.g., The battery ***can be used for*** almost an hour and can take as long as two hours to charge with an approximate distance of 30 meter.)	-

'SEE', 'FIND' AND 'NOTE' BUNDLES

Several other lexical verbs and their lexico-grammatical patterns have been recognised as the source of variations in the use of passive bundles in the BAWE sub-corpus in comparison to the MyPolyEELC. These include the 'see', 'find' and 'note' bundles as presented in Table 6. Similar to Hyland's (2008a) and Rezoug and Vincent's (2018) corpora, 'see' is found to occur within the BAWE sub-corpus. It appears that all the target passive bundles in relation to this lexical verb appear with a modal verb but with different lexico-grammatical patterns. The bundle, [12]*can be seen in* occurs very frequently throughout the corpus which carries the [12]'MODAL *be* V-ed *in* n' pattern with [12]'hear' as the meaning group which "is concerned with processes that take place in the mind, such as thinking, perceiving or feeling " (Francis et al., 1996, p. 50). Other lexico-grammatical patterns which display the same meaning group are

[13]*this* MODAL *be* V-ed prep (i.e., *this can be seen*), [14]*as* MODAL *be* V-ed prep (i.e., *as can be seen*) and MODAL *be* V-ed prep (i.e., *can be seen below*).

On the other hand, another 'see' bundle (i.e., [15]&[16]*can be seen to*) motivates two lexico-grammatical patterns (i.e., [15]MODAL *be* V-ed *to be* adj / [16]MODAL *be* V-ed *to be* V-ed) based on the language samples extracted from the BAWE sub-corpus. The meaning group which is related to these patterns is [15]&[16]'believe' which involves "thinking, saying or showing something" (Francis et al., 1996, p. 295).

The 'see' bundles also take [17]&[18]'think and discover' meaning group which is associated with the anticipatory *it* structural pattern. These bundles (i.e., [17]*it can be seen that* and [18]*it can be seen from*) which are patterned as '[17]*it* MODAL *be* V-ed *that*' and [18]'*it* MODAL *be* V-ed *from* n' are "concerned with what is thought" (Francis et al., 1996, p. 527). This includes the projection of the writer's opinion in relation to the subject of discussion.

Similarly, the 'find' bundles (i.e., [19]*it was found that*) and 'note' bundles (i.e., [20]*it should be noted*) which are realised as '[19]&[20]*it* MODAL *be* V-ed that' too take the same [17]-[20]'think and discover' meaning group. These similarities in terms of the lexico-grammatical pattern and the meaning group reveal that these forms of 'see', 'find' and 'note' bundles can be utilised interchangeably in a text. Both the present passive and past passive bundles of 'find' in the same pattern (i.e., *it is found that*; *it was found that*) are included in the list of 50 most frequently occurring 4-word bundles within the Electrical Engineering discipline (Hyland, 2008a, p. 12). 'Find' is also recorded in the list of 100 most frequent word families in the Student Engineering word list (Mudraya, 2006, p. 250). Another form of 'find' bundles which is prevalent in the BAWE sub-corpus is [21]*can be found in* (i.e., 'MODAL *be* V-ed *in* n') and carries a different meaning group (i.e., 'meet' which signifies "seeing or meeting someone or something in a particular place or situation") (Francis et al., 1996, p. 318).

TABLE 6. Meaning group associated with the lexico-grammatical patterns in both corpora for 'see', 'find' and 'note' passive bundles.

Meaning group	**Passive verb bundles**	**Lexico-grammatical pattern**	
		MyPolyEELC	**BAWE sub-corpus**
'hear' - concerned with processes that take place in the mind, such as thinking, perceiving or feeling	[12] can be seen in	-	[12] MODAL *be* V-ed *in* n (e.g., A clearer version of the graphs ***can be seen in*** Appendix A-C).
	[13] this can be seen	-	[13] *this* MODAL *be* V-ed prep (e.g., An example of ***this can be seen*** in the mode shapes of models 1A and 1B, where the mode shapes 2, 3 and 3, 2 occur opposite to one another).
	[14] as can be seen	-	[14] *as* MODAL *be* V-ed prep (e.g., ***As can be seen*** from figure 6 the hexagons will always appear slightly taller).
'think and discover' - concerned with what is thought	[17] it can be seen that	-	[17] *it* MODAL *be* V-ed *that* (e.g., However ***it can be seen that*** the hexagon provides a closer match for the circle than the square).

	[18] it can be seen from	-	[18] *it* MODAL *be* V-ed *from* n (e.g., ***It can be seen from*** these results that the stress distribution is no longer even, but varies with the depth of the section).
	[19] it was found that	-	[19] *it* MODAL *be* V-ed *that* (e.g., ***It was found that*** using a ruler to measure the diameters would be inaccurate and imprecise, as the ruler would have to be replaced on the fluorescent screen every time a new measurement was required).
	[20] it should be noted	-	[20] *it* MODAL *be* V-ed *that* (e.g., ***It should be noted*** **that** where only three parts to an experiment exist, a strong correlation is not as persuasive evidence for consistency as when many parts can be compared).
'believe' - thinking, saying or showing something	[15] & [16] can be seen to	-	[15] MODAL *be* V-ed *to be* (e.g., From the graph this material ***can be seen to*** be stiffer, stronger and more ductile than the other two materials). [16] MODAL *be* V-ed to-inf (e.g., In the schematic figure B.9, an output pin labelled 'LEDG' ***can be seen to*** be connected to Vcc for this purpose).
'meet' - seeing or meeting someone or something in a particular place or situation	[21] can be found in	-	[21] MODAL *be* V-ed *in* n (e.g., The hard copies of the waveforms produced ***can be found in*** the graph section).

FUNCTIONAL CATEGORIES

SIGNIFICANCE OF DIFFERENCES

The log-likelihood test was performed on the raw frequencies of the functional categories identified between the L1 and L2 learner corpora. It is important to see if the differences between the corpora show significant likelihood value for further investigation to discuss the pedagogical implications. The analysis showed a significant difference in the frequency of functional categories. The results indicated log-likelihood (LL) value above 15.13 for 'procedure', 'description', 'resultative signals', 'engagement' and 'stance' functional categories which was significant at the level of $p<0.0001$ with the Bayes Factor (BIC) measure score of >10 which established very strong evidence that there was indeed significant

difference in the use of these functional categories between the L1 and L2 learner corpora (see Table 7).

TABLE 7. The log-likelihood test results

Functional Categories	MyPolyEELC (Raw frequency)	BAWE (Raw frequency)	Likelihood value (LL)	Bayes Factor (BIC)
Procedure	851	163	197.17	183.80
Description	119	19	34.24	20.87
Engagement	74	219	172.28	158.91
Stance	0	29	59.38	46.01
Resultative Signals	0	61	67.57	54.20

'PROCEDURE' AND 'DESCRIPTION' BUNDLES

In line with the nature of the texts involved in this study, (i.e., technical academic writing) the research-oriented category (i.e., 'procedure' and 'description') has the highest number of occurrences in the L2 learner corpus. However, the L1 learner writing shows a greater use of participant-oriented category as the most frequently occurring functional category across the corpus (see Table 9 in Appendix). The research-oriented [22], [23], [25] & [26] 'procedure' bundles that indicate the core activities or rather steps in the project work and [24]'description' bundles which further elaborate or define the components and features in the project frequently occur in the form of [22]-[23]'passive verb + PP fragment' structure in both MyPolyEELC and BAWE corpus. Apart from the 'passive verb + PP fragment', the 'procedure' bundles in the BAWE sub-corpus are almost equally associated with the [26]'verb phrase + to-clause' pattern unlike the MyPolyEELC which shows much higher use of 'procedure' bundles in 'passive verb + PP fragment' pattern. 'Use' is identified as a procedural verb (Lim, 2017) which is similar to the findings of Rezoug and Vincent (2018, p. 72) who have highlighted several bundle clusters including the verb 'use' to "describe procedures" as in the examples like "*can be used to/for/in* and *'it/which is used to*". Biber et al. (1999, p. 382) categorise 'use' as a type of "activity verb". The 'use' bundles in the present study are treated as 'procedure' bundles which demonstrate the steps implemented or activities performed throughout the project (Durrant, 2017; Hyland, 2008a; Hyland, 2008b) although the modal auxiliary *can* is often associated with stance markers which indicate possibility or ability (Biber, 2006; Biber et al., 2004). In addition, 'use' is treated as 'procedure' as the subject often appears as the "instrument" utilised by the "scientist or "technician" to execute certain activity (Durrant, 2017, p. 185). Table 8 shows instances of the meaning group associated with the functional categories in both corpora.

'ENGAGEMENT' AND 'STANCE' BUNDLES

There are 6 participant-oriented categories traced in the form of anticipatory *it* in the L1 native speakers' writing in the BAWE sub-corpus with no traces of this type in the L2 learner writing. As contended by Hyland (2008a), the anticipatory *it* structure is associated more closely to participant-oriented category. However, the use of this category throughout the BAWE sub-corpus is not only limited to [29] - [31]anticipatory *it* pattern but 'engagement' has been utilised extensively in [27],[32]&[37]'passive verb+ PP fragment', [28]'adverbial clause fragment' and [36]'verb phrase + *to*-clause'. The number of participant-oriented categories identified in the BAWE sub-corpus is 17 altogether with 14 engagement markers and 3 stance markers. The L2 learner writing, on the other hand illustrates 8 engagement features (e.g., [34] – [35]) with no stance category identified. This shows fewer use of participant-oriented functional category among the L2 writers compared to L1 English writers. This finding correlates with Chen and

Baker's (2010) study which identifies higher use of 'stance' category among the BAWE L1 English learners (i.e., 24%) in contrary to BAWE Chinese learners (L2 learners) (i.e., 16%). 'Stance' expression involves a writer's stand towards the discussion. Its delivery in the form of writing requires advanced academic writing skills. The influence of the second language is also identified as one the factors which results in the limited use of the category in the writings of the L2 learners (Hyland, 2008a). Despite the presence of participant-oriented category in L2 learner writing, the lexico-grammatical pattern which indicates the choice of word reveals a repeated use of modal verb *can* that is categorised as an attitudinal stance marker (Biber et al., 2004). The present analysis shows that this stance marker is however more apt to be categorised within the engagement feature as it demonstrates the writer's interaction with the reader by minimising the imposition on the reader as the modal verb *can* depicts the possibility / ability (Biber, 2006).

The analysis in the present study indicates six lexical verbs (i.e, [27],[28]&[36]*see*, [33],[37],[38]&[40]*find*, [29]*decide*, [30]*assume*, [31]*note* and [32]*take*) conveying the 'engagement' and 'stance' discourse functions (from the L1 learner writing) which however, have not been found in the form of bundles utilised by the L2 learners. These bundles which are only prevalent in the native speaker learner writing can further be considered for pedagogical purposes to enhance the L2 learners' technical writing to directly highlight their writer/reader-oriented involvement in the text (see Table 8).

'RESULTATIVE SIGNALS' BUNDLES

The presence of 'resultative signals' can be seen only in the target passive bundles analysed in the BAWE sub-corpus. Discussing the results and discovery from the analysis, this text-oriented category emerges in [38]&[39]'verb phrase + to-clause' and [40]'anticipatory *it* + verb phrase' structural patterns. Rezoug and Vincent (2018) too identify this discourse function to be infrequent in their L2 learner corpus. The highest number of occurrences of the research-oriented bundles in the L2 learner corpus shows that these learners are inclined to describing the research activities instead of reporting or discussing their findings (see Table 8).

TABLE 8. Meaning group associated with the functional categories in both corpora

Meaning group	**Functional Categories**	**Passive verb bundles / Structural patterns**	
		MyPolyEELC	**BAWE sub-corpus**
'use' -the role assigned to something in the course of a particular action	[22] & [23] RO: Procedure	[22] can be used as [passive verb + PP fragment] (e.g., Some embedded projects also consists of IR Transmitter and Receiver Modules where they ***can be used as*** proximity sensors or distance measurement sensors).	[23] is applied to the [passive verb + PP fragment] (e.g., When a force ***is applied to the*** beam it causes the strain gauges that are attached to the beam to undergo either compressive or tensile forces).
'discover' -concerned with coming to know something or bringing something to mind	[37] PO: Engagement	-	[37] can be found by [passive verb + PP fragment] (e.g., Having identified the first resonances, subsequent resonances ***can be found by*** matching them to this line).

	[38] TO: Resultative Signal	-	[38] is found to be [verb phrase + *to*-clause] (e.g., The floating caliper system ***is found to be*** far less sensitive to high temperatures, and consequently, this type of brake failure is uncommon).
'provide' -giving something to someone or something	[24] RO: Description	[24] is equipped with sets of [passive verb + PP fragment] (e.g., The board ***is equipped with sets of*** digital and analog input/output pins that may be interfaced to various expansion boards and other circuits).	-
'choose and use' - appointing or choosing someone to do something or with allocating or assigning something to a particular use	[25] & [26] RO: Procedure	[25] can be used to [verb phrase + *to*-clause] (e.g., Arduino ***can be used to*** communicate with a computer, another Arduino board or other microcontrollers).	-
		-	[26] will be used to [verb phrase + *to*-clause] (e.g., Strain gauge readings attached to the beams surface ***will be used to*** monitor the strains in the beam as the load is applied).
'believe' - thinking, saying or showing something	[36] PO: Engagement	-	[36] can be seen to [verb phrase + *to*-clause] (e.g., It ***can be seen to*** match the requirement shown in figure 3.5.1).
	[39] TO: Resultative Signal	-	[39] is said to be [verb phrase + *to*-clause] (e.g., If the velocity profile across the jet decreases smoothly to zero, the flow in the jet ***is said to be*** laminar).
'hear' - concerned with processes that take place in the mind, such as thinking, perceiving or feeling	[27] & [28] PO: Engagement	-	[27] can be seen in [passive verb + PP fragment] (e.g., For each specified engine speed the engine torque and power were recorded and ***can be seen in*** Table 1). [28] as can be seen [adverbial clause fragment] (e.g., ***As can be seen*** from this brief description of the CAD model geometry, it is advantageous to plan the modelling of the geometry construction so that additional

			features can be added that make the analysis process easier).
'decide' **-** thinking or talking about how to do something or whether to do something	[29] PO: Stance	-	[29] it was decided that [Anticipatory *it* + verb phrase] (e.g., ***It was decided that*** rather than split up responsibilities to each team member it would be more time efficient to work on a single problem collectively).
'think' **-** concerned with thinking, including understanding something	[30] PO: Stance	-	[30] it is assumed that [Anticipatory *it* + verb phrase] (e.g., ***It is assumed that*** the pressure readings are the same for the bottom surface of the cylinder as the top pressure readings).
'involve' **-** concerned with making someone become involved in an activity	[32] PO: Engagement	-	[32] must be taken into [passive verb + PP fragment] (e.g., In calculating the route that the spacecraft will take, a number of things ***must be taken into*** consideration).
'think and discover' - concerned with what is thought	[40] TO: Resultative Signal	-	[40] it was found that [Anticipatory *it* + verb phrase] (e.g., Designing for the pinion gear, ***it was found that*** using the same dimensions as a metal gear used in the program would produce a gear that would not support the stress loads placed upon it).
	[31] PO: Engagement	-	[31] it should be noted [Anticipatory *it* + verb phrase] (e.g., ***It should be noted*** that the theoretical calculations used to determine the yield load of the I section beam use a much simplified model compared to the actual component).
'meet' - seeing or meeting someone or something in a particular place or situation	[33] PO: Engagement	-	[33] can be found in [passive verb + PP fragment] (e.g., Results of the convergence testing ***can be found in*** the results section of this report, along with the actual analysis results).

'fight' - concerned with moving somewhere with force or with difficulty, either physically or metaphorically. This includes moving with energy	[34] PO: Engagement	[34] it can be powered by [Anticipatory *it* + verb phrase] (e.g., ***It can be powered by*** a USB cable or by an external 9 volt battery, though it accepts voltages between 7 and 20 volts).	-
'end' -bringing a situation to an end	[35] PO: Engagement	[35] can be completed in [passive verb + PP fragment] (e.g., A step by step procedure is done so that the project ***can be completed in*** time).	-

CONCLUSION AND RECOMMENDATIONS

This study is an attempt to reveal the lexico-grammatical patterns and functional categories of passive verb bundles between native and non-native learner writing. Regardless of the functional categories, the 'passive + PP fragment' structural pattern is found prominent in both the L1 and L2 learner corpora. This finding coincides with Hyland's study (2008a, p. 11) where he identifies passive verb with prepositional phrase combination as the most frequently occurring in the hard sciences to "downplay the personal role" of the researcher/scientist to establish that despite anyone conducting the study, the results remain unchanged.

In contrast to the MyPolyEELC, the L1 learners in the BAWE sub-corpus attempt more uses of participant-oriented 'engagement' and 'stance' features (e.g., *it should be noted*, *it is assumed that*). These discourse functions include the reader-writer relationship and the stance bundles which express the writer's involvement in the text to present voice/opinion, require more advanced academic writing which is perhaps less challenging for the native speaker writers (Hyland, 2008a). This result is similar to that of several studies which have identified more uses of stance-oriented bundles among the native speaker learners compared to non-native speaker learners (Adel & Erman, 2012; Chen & Baker, 2010). Likewise, Pan and Liu (2019) identify that L1 student and expert writers tend to employ stance bundles to highlight the writer's element of judgement and evaluation towards a proposition instead of merely introducing visuals or tables which are presented in the text.

The loglikelihood significance findings of the study suggest that the use of 'engagement' and 'stance' bundles in the writings of a technical report together with text-oriented 'resultative signals' to introduce results of the project or research, needs to be emphasised pedagogically, among the L2 learners. How these bundles are utilised within the specific texts by looking at the lexico-grammatical patterns and associated meaning groups need to be identified, for instance, '*see*', '*find*', '*note*', '*take*', '*assume*' and '*decide*' bundles as discovered from the L1 English learners' use in the BAWE sub-corpus to further devise the teaching and learning activities through data-driven learning techniques for the L2 learners to enhance their technical report writings in terms of passive verb bundles' use. Variations in the word choice and patterns can also be instructed to these learners as different lexical verbs but carrying the same lexico-grammatical pattern and associated meaning can be used interchangeably. This is important as the findings of the present study show repeated

occurrence of the '*use*' bundles as the major option for the L2 learners to rely on when describing the procedures.

The findings also reveal the inter-relation between the meaning group, lexico-grammatical pattern and functional category where the meaning group which motivates the lexico-grammatical pattern further motivates the functional category of the bundles (e.g., *it was found that*, *can be found in* and *can be found by*). Additionally, ill-formed passive bundles too shall not be neglected but addressed pedagogically as the analyses of this study also discovered some misused target passive bundles (which is not discussed in this study) by the L2 learners which may be associated with verb-related errors produced by Malaysian learners (e.g., Ang et al., 2020). One of the limitations of the current study could be that the gap identified in the L2 learner writing is confined to the analysis of passive verb bundles with no other types of structural categories that may demonstrate participant-oriented functions are included. Future studies may further broaden the focus by identifying the lexico-grammatical pattern and functional category of the verb phrase-based bundles which may reveal more structural types associated with the pattern and meaning to address the L2 writers' academic writing.

ACKNOWLEDGEMENT

This PhD study is sponsored by the Ministry of Education, Malaysia, under the Federal Training Scholarship.

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. https://doi.org/10.1016/j.esp.2011.08.004

Ang, L. H., Tan, K. H., & Lye, G. Y. (2020). Error types in Malaysian lower secondary school student writing: A corpus-informed analysis of subject-verb agreement and copula *be*. *3L: The Southeast Asian Journal of English Language Studies*, *26*(4), 127-140.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam: John Benjamins Publishing Company.

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

Biber, D., & Conrad, S. (2019). *Register, genre and style*. (2nd ed.). UK: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at …..* Lexical bundles in University teaching and textbooks. *Applied Linguistics*, *25*(3), 371-405.

Biber, D., & Grey, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL Ibt test: A lexico-grammatical analysis. *TOEFL Ibt -19 Research Report*. ETS. https://doi.org/10.1002/j.2333-8504.2013.tb02311.x

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

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, *14*(2), 30-49.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165–193. https://doi.org/10.1093/applin/amv011

Firth, J. R. (1957). *Papers in linguistics 1934-1951*. Oxford: Oxford University Press.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 137–56). Amsterdam, Netherlands: John Benjamins.

Francis, G., Hunston, S., & Manning, E. (1996). *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins Publishers Ltd.

Granger, S. (2013). The passive in learner English: Corpus insights and implications for pedagogical grammar. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1,* 5-15. Papers from LCSAW2013. Kobe: School of Languages and Communication, Kobe University.

Green, C. (2019). Enriching the academic wordlist and Secondary Vocabulary Lists with lexicogrammar: Toward a pattern grammar of academic vocabulary. *System*, *87*, 1-10. https://doi.org/10.1016/j.system.2019.102158

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In Granger, S., & Meunier, F. (Eds.) *Phraseology*: *An interdisciplinary perspective* (pp. 3-25). Amsterdam: John Benjamins Publishing Company.

Hinkel, E. (2002). Why English passive is difficult to teach (and learn). In Hinkel, E., & Fotos. S. (Eds.), *New perspectives on grammar teaching* (pp. 233-260). New York: Lawrence Erlbaum Associates.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Bonelli, E. T. (ed). The Netherlands: John Benjamins Publishing Company.

Hussain, G., Zahra, T., & Abbas, A. (2021). Discourse functions of lexical bundles in Pakistani Chemistry and Physics books. *GEMA Online Journal of Language Studies*, *21*(1), 221-238.

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

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 4-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., et al. (2014). The Sketch Engine: ten years on. *Lexicography ASIALEX*, *1*, 7-36. DOI 10.1007/s40607-014-0009-9.

Lee, Y. E., Yoo, I. W., Shin, Y. K. (2020). The use of English prepositions in lexical bundles in essays written by Korean university students. *Journal of English for Academic Purposes*, *45*, 1-10. https://doi.org/10.1016/j.jeap.2020.100848

Liardet, C.L., & Black, S. (2019). "So and so" says, states and argues: A corpus-assisted engagement analysis of reporting verbs. *Journal of Second Language Writing. 44*, 37-50.

Lim, J. M. H. (2017). Writing descriptions of experimental procedures in language education: Implications for the teaching of English for academic purposes. *English for Specific Purposes*, *47*, 61-80. https://doi.org/10.1016/j.esp.2017.05.001

McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus Linguistics, Learner Corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics*, *39*, 74–92. https://doi.org/10.1017/s0267190519000096

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*(2), 235–256. https://doi.org/10.1016/j.esp.2005.05.002.

Otto, P. (2021). Choosing specialized vocabulary to teach with data-driven learning: An example from civil engineering. *English for Specific Purposes. 61*, 32-46. https://doi.org/10.1016/j.esp.2020.08.003

Pan, F., & Liu, C. (2019). Comparing L1-L2 differences in lexical bundles in student and expert writing. *Southern African Linguistics and Applied Language Studies*, *37*(2), 142-157.

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

Qian, M., & Ma, H. (2020). The creation and evaluation of a grammar pattern list for the most frequent academic verbs. *English for Specific Purposes*, *58*, 155-169. https://doi.org/10.1016/j.esp.2020.01.002

Quirk, G., Greenbaum, S., Leech, G., Svartvik, J. (1985). *A comprehensive grammar of the English language*. USA: Longman.

Rayson, P. (2006). Log-likelihood calculator [Computer software]. Lancaster University. Available from http://ucrel.lancs.ac.uk/llwizard.html

Rezoug, F., & Vincent, B. (2018). Exploring lexical bundles in the Algerian Corpus of Engineering. *Arab Journal of Applied Linguistics*, *3*(1), 47–77.

Roslina Abdul Aziz. (2018). A corpus-based study of the use of *"BE"* in Malay ESL learner essays. Unpublished PhD Thesis, Universiti Malaya.

Scott, M. (1997). PC analysis of keywords – and key key words. *System*, *25*(2), 233-245.

Sinclair, J. (1996) The search for units of meaning. *Textus*, *9*(1), 75–106.

Sinclair, J. (2005). Corpus and Text – Basic Principles. In Wynne, M. (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxbow Books.

Sinclair, J. (2009). The phrase, the whole phrase, and nothing but the phrase. In Granger, S., & Meunier, F. (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 407-410). Amsterdam: John Benjamins Publishing Company.

Siti Aeisha Joharry. (2016). Malaysian learners' argumentative writing in English: A contrastive, corpus-driven study. Unpublished PhD Thesis, University of Sydney.

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres and language ratings. *Journal of Second Language Writing*, *32*, 17-35. https://doi.org/10.1016/j.jslw.2016.02.002

Stubbs, M. (2009). Technology and phraseology: With notes on the history of corpus linguistics. In Romer, U & Schulze, R. (Eds.), *Exploring the lexis-grammar interface* (pp.15-31). Amsterdam: John Benjamins Publishing Company.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (2nd ed.). Ann Arbor: The University of Michigan Press.

Teh, H. B. L., Zuraida Yaacob, Aziam Mustafa & Mohd Hasni Angterian. (2016). Writing format. In Siti Noridah Ali & Alinawate Ali. (Eds.), *Diploma students' project guide* (pp. 28-78). Putrajaya: Instructional and Digital Learning Unit, Department of Polytechnic and Community College Education.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Company.

eISSN: 2550-2131
ISSN: 1675-8021

APPENDIX

TABLE 9. List of finalised passive verb bundles with functional category in the MyPolyEELC and BAWE sub-corpus

No.	MyPolyEELC			BAWE		
	Item	Raw Frequency	Functional Category	Item	Raw Frequency	Functional Category
1	can be used to	78	RO:P	can be seen in	92	PO:E
2	can be used as	70	RO:P	can be used to	34	RO:P
3	used in this project	61	RO:P	it can be seen that	31	PO:E
4	as mention in previous chapter	51	TO:SS	was found to be	28	TO:RS
5	related to the project	42	TO:FS	as shown in figure	20	TO:SS
6	is connected to the	36	RO:P	it was found that	17	TO:RS
7	that can be used	32	RO:P	can be found in	17	PO:E
8	obtained from the proposed methodology	30	TO:SS	will be used to	15	RO:P
9	can be used for	30	RO:P	is shown in figure	14	TO:SS
10	needed to support the microcontroller	29	TO:FS	used to determine the	12	RO:P
11	may be interfaced to various	29	RO:D	it was decided to	12	PO:S
12	will be sent to	28	RO:P	is given by the	12	TO:SS
13	based on the atmega328	25	TO:FS	shown below in figure	11	TO:SS
14	is realize using proteus software	25	RO:P	it was decided that	10	PO:S
15	equipped with sets of digital	25	RO:D	can be seen to	10	PO:E
16	as shown in figure	23	TO:SS	as can be seen	10	PO:E
17	displayed on the lcd	23	RO:P	was calculated to be	9	RO:P
18	will be used to	23	RO:P	was applied to the	9	RO:P
19	is based on the	22	TO:FS	can be applied to	9	PO:E
20	can also be used	22	RO:P	used to measure the	8	RO:P
21	used to detect the	21	RO:P	is said to be	8	TO:RS
22	can be used in	20	RO:P	is found to be	8	TO:RS
23	it can be used	19	RO:P	is applied to the	8	RO:P
24	it is used to	19	RO:P	be used to calculate	8	RO:P
25	used to write and upload	19	RO:P	be added to the	8	RO:P
26	made in this project	19	RO:P	to be used in	7	RO:P
27	can be completed in	18	PO:E	to be carried out	7	RO:P
28	done so that the project	17	RO:P	must be taken into	7	PO:E
29	it is partitioned into	17	TO:SS	it is assumed that	7	PO:S
30	related to this project	15	TO:FS	is defined as the	7	RO:D
31	needed to complete it	13	TO:FS	could be used to	7	RO:P
32	used to control the	13	RO:P	be used to produce	7	RO:P
33	is described in this chapter	13	TO:SS	it can be seen from	7	PO:E

eISSN: 2550-2131
ISSN: 1675-8021

34	designed for transparent	12	RO:D	used to find the	6	RO:P
35	wireless serial is focused especially on	12	TO:FS	this can be seen	6	PO:E
36	it can be powered by	12	PO:E	that can be used	6	RO:P
38	will be employed in	11	TO:SS	it should be noted	6	PO:E
39	as shown in the	11	TO:SS	it can be shown	6	TO:SS
40	referred to as a	11	RO:D	is known as the	6	RO:D
41	needs to be achieved	11	TO:FS	is assumed to be	6	RO:P
42	is designed by using	11	RO:P	can be seen below	6	PO:E
43	was used in this	11	RO:P	can be found by	6	PO:E
44	created using components available	11	RO:P	can be expressed as	6	PO:E
45	have been conducted in	10	RO:P	can be defined as	6	PO:E
46	is discovered about the	10	TO:SS	be calculated from the	6	RO:P
47	that will be used	10	RO:P	referred to as the	6	RO:D
48	is applied to the	10	RO:P			
49	will be stored in	10	RO:P			
50	obtained in this chapter	9	TO:FS			
51	that have been done	9	RO:P			
52	needs to be done	9	PO:E			
53	related to the research	9	TO:FS			
54	are also used for	9	RO:P			
55	is used as the	9	RO:P			
56	that have been used	9	RO:P			
57	that used in this	9	RO:P			
58	used as an input	9	RO:P			
59	was chosen to mark	9	RO:D			
60	can be controlled by	9	PO:E			
61	obtained from the experiments	8	TO:FS			
62	employed by a discipline	8	TO:FS			
63	is shown in figure	8	TO:SS			
64	can be connected to	8	RO:P			
65	used for this project	8	RO:P			
66	been used and others component	8	RO:P			
67	used to make this	8	RO:P			
68	discussed in this chapter	8	TO:SS			
69	that is written in the	8	RO:D			
70	can be done by	7	PO:E			

71	done to see the	7	RO:P
72	it is designed to	7	RO:P
73	that have been designed	7	TO:FS
74	will be connected to	7	RO:P
75	used for loading programs	7	RO:P
76	used in various devices	7	RO:P
77	is used for the	7	RO:P
78	that can be *use*	7	RO:P
79	used for building electronics projects	7	RO:P
80	used in the system	7	RO:P
81	used to add a reset	7	RO:P
82	used to measure the	7	RO:P
83	carried out throughout the	7	RO:P
84	are soldered to the	7	RO:P
85	are pressed into pcb board	7	RO:P
86	can be configured to	7	RO:P
87	can be detected by	7	PO:E
88	is organised into five	7	TO:SS
89	is released under the	7	RO:D
90	it is connected to	6	RO:P
91	can be used with	6	RO:P
92	are divided into two	6	RO:D
93	can be installed in	6	PO:E
94	linked to the rules	6	RO:D
95	is distributed under a	6	RO:D
96	should be arranged as	6	PO:E

RO:P – Research-oriented: Procedure, RO:D – Research-oriented: Description, RO:T – Research-oriented: Topic, RO:L – Research-oriented: Location, RO:Q – Research-oriented: Quantification, TO:SS – Text-oriented: Structuring signal, TO:FS – Text-oriented: Framing signals, TO:RS – Text-oriented: Resultative signals, PO:E – Participant-oriented: Engagement, and PO: S – Participant-oriented: Stance

ABOUT THE AUTHORS

Radika Subramaniam is a PhD candidate at the Faculty of Languages and Linguistics, Universiti Malaya. Her research interests include integrating corpus linguistics into ESP and EAP studies, particularly the academic written discourse.

Sheena Kaur is a Senior Lecturer and currently the Deputy Dean of Postgraduate Studies at the Faculty of Languages and Linguistics, Universiti Malaya. She obtained her PhD in Applied Linguistics from Lancaster University, UK. Her main research interests include Corpus Linguistics (i.e., application of corpus-based methods in semantic and text analysis), applied linguistics, higher education and Covid-19 related studies.

eISSN: 2550-2131
ISSN: 1675-8021