Lexical bundles in engineering research articles: materials development for Thai engineering students

Kamonchanok Sanmuang

BA in English for International Communication, Rajamangala University of Technology, Isan, Thailand

MA in Applied Linguistics, King Mongkut's University of Technology, Thonburi, Thailand

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Applied Linguistics

Department of Linguistics and English Language

Lancaster University

March 2018

Declaration

I hereby declare that this thesis is my own work, and has not been submitted in substantially the same form for the award of a higher degree elsewhere.

Kamonchanok Sanmuang

March 2018

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my supervisor, Dr Karin Tusting, who has provided me with insightful comments and emotional support throughout my PhD adventure. Without her guidance and persistent help, I would never have completed this thesis. Thank you for your patience and constant support.

I would like to thank Dr Jane Sunderland, our supportive Director of Studies, who founded the T&CW programme in our department. Thank you, Jane, for your help during the residential stays and for your helpful suggestions in pursuing this PhD.
I am also indebted to all the staff members of the Linguistics Department, especially to our Head of Department, Professor Elena Semino, for keeping our department on top form. I would also like to express my sincere thanks to our very well-organised Coordinators, Marjorie Wood and Elaine Heron, for their support throughout these years.
I would like to express my appreciation to Professor Greg Myers, Professor Paul Baker, Professor Yukio Tono, and Dr Vaclav Brezina for their valuable suggestions throughout different stages of this thesis. Sincere thanks to Gerard Hearne for doing a fantastic job in proofreading throughout various stages of my PhD and this thesis.

In addition, I would like to thank all my friends and colleagues at Kasetsart University, Chalermphrakiat Sakon Nakhon Campus (KU.CSC). In particular, I would like to thank Dr Watcharapong Intrawong, a previous Dean of Faculty of Liberal Arts and Management Science, and Dr Siriluck Prasanpangsri, Dean of Faculty of Liberal Arts and Management Science, who gave me an invaluable opportunity to pursue this degree in the UK. I also thank the Vice President of KU.CSC, who has sponsored me throughout the PhD study. I also thank the Dean of Faculty of Engineering and all the staff who supported and were involved in the 'lexical bundles workshop'. In particular, I would like to thank the group of Electrical and Computer Engineering students who were patient and attended all the sessions of the lexical bundles workshop; without their cooperation, I would not have had such great data.

I would like to thank all my friends and colleagues, Assistant Professor Dr Chalard Jukpim (Pee Pik), Dr Techatouch Kaisoka (Pee Yul), Dr Piyanoot Ruangpon (Pee Peuang), and Dr Oruma Taeporamesamai (Pee Ying), for helping me with all the scholarship document processes in Thailand, and for being good friends and colleagues at the same time, since I started working at KU.CSC. Thank you to any other people at KU.CSC whom I have not mentioned here.

I am grateful for the ongoing support from my brother and friends, Pee Pik, Associate Professor Dr Atipat Boonmoh, who have encouraged me not only academically but also in ways that taught me how to think positively and work hard and smart. I am also deeply thankful to Nong Parn, Dr Natkristha Kophimai who has cheered me up throughout this journey. My heartfelt thanks also go to Pee Tom, Dr Taweewat Inree and Pee Pulom, Associate Professor Dr Nucharee Nuchkoom Smith, whom I have known for years as a colleague. Your support and encouragement have meant so much to me during this time.

Thank you to all my Lancaster friends who are not involved in academia, but whose valuable help in living life were tremendously useful: Pee Thom, Pee Big, Pee Porn, Pee Yom, Pee Nuch, and Kevin.

I would like to thank my dear friends Pee Ying, Pee Jacky, Chad, Fon, Kwang, Nong, Mui, Bew, Num, Kriangkrai, Tahir, Pamela, João, Hsiao-Yun, Sally, Stanley, Nick, Maria-Elena, and many others who were there for me during good times and bad times. Thank you for your support and encouragement.

I owe a big thanks to my very knowledgeable and kind friend, Dr Porntip (Mameaw), for her invaluable comments and suggestions at different stages of the PhD thesis; she is like my sister and my friend. I learnt a lot from her. Thank you for your constant support, both academically and emotionally.

Extraordinary thanks go to my lovely and genuine friend, Zoe. Your encouragement inspired me a lot to keep going with my PhD. Your positive vibes were all around me, and I will remember the activities we did together, apart from the PhD: chatting, having dinner, and working out at the gym. I will never forget what we experienced together and all the hurdles we overcame. We were always there to support each other.

This thesis would never have been completed without Samson, my husband. He is like my best friend and my brother at the same time. Another person who reminds me of real moments in life is Chris, my son of two years and six months. He always gives me strength whenever I am weak; he is my oxygen. It was a very tough time when I had to leave Chris when he was eight months old. I will never forget the time when I was continuously crying on the plane, on the way back from Thailand to the UK. It is painful to say that Samson and Chris had to live without me in Thailand while I was writing up my thesis in Lancaster. Even though I have missed the chance of seeing his recent development stages—walking, starting to say words—it will be worth it in the end. Metaphorically speaking, I wanted to stay focused on finishing the thesis chapters. Pee Lew, my husband, has always been there for me to support me emotionally. I am grateful to have both of them. I cannot wait to see them again!

Last but not least, I want to express my gratitude to my family: my mother-in-law and my father-in-law, my mother and my father. Even though my dad passed away, I think I have made him proud of me. This thesis would never have been started or completed without them. Thank you for teaching me how tough life is, how to be patient, and never give up, despite everything that happens in life. I am indebted to them for making big sacrifices and taking good care of my son, Chris, to give me time in which to complete the thesis in Lancaster. They have always supported me emotionally and financially.

Abstract

The aim of this study is to investigate the potential and impact of using analyses of lexical bundles to support Thai engineering students in learning to write research articles, as well as to develop a sample of materials for engineering students to use when writing research articles.

This research focuses on 200 3-word and 4-word bundles extracted from a 400,000-word corpus compiled from a collection of 100 research articles on engineering. The Top 50 and Bottom 50 3-word and 4-word bundles were analysed to identify their structures and functions. Frequency and dispersion range were used as the main criteria in lexical bundle identification (Conrad & Biber, 2005). Additionally, inter-rater reliability was used to enhance validity and reliability.

The study uses the structural and functional classification of lexical bundles in the Longman Grammar of Spoken and Written English from Biber, Finegan et al. (1999). The structures of the top 50 lexical bundles were categorised into Noun Phrase-based (NP-based), Preposition Phrase-based (PP-based), Verb Phrase-based (VP-based), Clausal fragments and other expressions. Preposition Phrase-based (33%), VP-based (28%) and NP-based (26%) appeared to be the most frequent lexical bundles. Other expressions (that did not fall into the previous categorisation) (7%) and clausal fragments (6%) occurred less than the three groups mentioned. Regarding the structures of the less frequent lexical bundles (the bottom 50), the study found that bundles were in three main structural categories: incomplete noun phrase fragments (65%), incomplete dependent clauses (30%), and other expressions (5%).

Regarding discourse functions, referential expressions (49%) and discourse organisers (34%) of the top 50 lexical bundles were more frequent than stance (16%) and others (1%). The study employs structural and classification results to formulate/frame the materials. The results of the bottom 50 revealed similar proportions to the top 50: referential (39%), discourse organisers (29%), others (27%), and stance (5%). The inter-rater reliability of the structural categorisation of the top 50 and bottom 50 was 97%. The inter-rater reliability of the functional classification of the top 50 and bottom 50 was also 97%.

For instructional purposes, five of these 200 bundles from the top 50 were selected for teaching 13 electrical engineering students. With the aid of AntConc software, the students worked with lexical bundles at their own pace, with help from the researcher whenever necessary. A pre-test and a post-test along with student interviews, students' notes, classroom observation and a teacher's diary were used as the main research instruments to assess the effectiveness of the process of using materials for using lexical bundles developed in the study and classroom instruction in lexical bundles.

The post-test scores for the writing task were significantly higher than those in a pre-test. The results demonstrated that the students showed a more positive attitude towards learning English in general and revealed more awareness regarding their use of lexical bundles employed in the engineering field.

The findings from the lexical bundle analyses shed light on the methodology for analysing lexical bundles, which pays more attention to not only frequent lexical bundles but also less frequent lexical bundles that are more relevant to the subject domain. Pedagogically, the experiment suggested that there is a need to incorporate lexical bundles and data-driven learning in teaching subject-specific writing. This study could be replicated with other groups of students in subject-specific areas.

List of Abbreviations for Lexical Bundles

LBs	Lexical bundles
ERAs	Engineering research articles
ERAs	Corpus Engineering research articles corpus
ERAC	Engineering research articles corpus
LGSWE	Longman grammar of spoken and written English
Noun Phrase	NP
Preposition Phrase	PP
Verb Phrase	VP
Clausal fragment	CF

Other Abbreviations

§	see section
ESP	English for Specific Purposes
EAP	English for Academic Purposes
KU.CSC	Kasetsart University, Chalermphrakiat Sakon Nakhon Province Campus

Table of Contents

List of Figures

List of Tables

1. Chapter One: Introduction

1.1. Background, motivation, and rationale

In corpus linguistics, corpus-based studies typically involve using corpora to test hypotheses or explore linguistic phenomena, whereas corpus-driven studies rely on corpora to generate hypotheses and theories (McEnery & Hardie, 2011). The prime objective of this study is to investigate the impact of using analyses of lexical bundles to support Thai engineering students in learning to write research articles, as well as to develop a sample of corpus-driven materials for use by engineering students when writing research articles. This study follows a corpus-driven approach, which involves identifying patterns and features directly from the data, rather than relying on pre-existing linguistic categories. In contrast, corpus-based studies are typically guided by pre-established linguistic frameworks. This research relies on methods of analysing formulaic languages, such as frequency and dispersion range, to be able to analyse engineering research articles (henceforth, ERAs) (cf. Biber et al., 1999; Conrad & Biber, 2005), see examples in Chapter 3. Drawing on the lexical bundle analyses from Chapter 3, this research focuses on corpus techniques, such as corpus-driven learning, and other strategies for teaching the structure and function of lexical bundles to engineering students in the classroom (cf. John, 1991; McEnery, Baker, & Wilson, 1995; McEnery & Gabrielatos, 2008). For example, structurally, lexical bundles can incorporate noun phrase (NP) and prepositional phrase (PP) fragments, such as *one of the* (*100*) and *the effect of the* (*45*), which are used to identify or elaborate on physical or abstract entities. Verb phrase (VP)-based bundles, like *it can be* (*171*) and *it can be seen* (*57*), often express possibilities, actions, or results. Additionally, dependent clause fragments, such as *to determine the* (*73*), serve to introduce purposes or specific actions within engineering discourse. Functionally, these bundles may serve referential purposes, discourse organisation, or stance marking, helping students to understand and produce clear, concise, and well-structured academic and technical writing. Further applications and examples of these bundles are discussed in Chapter 4.

Corpus-based studies offer many benefits to language description and language teaching in academic discourse (cf. Leech, 1997; Tognini-Bonelli, 2001; Hyland, 2015).

Individually, corpus linguistic techniques make it possible for linguists to identify and classify linguistic features and discourse functions in academic corpora (Conrad, 2015). Corpus-based methods can allow us to investigate the characteristics of texts on the macro and micro levels. For example, linguistic features and discourse functions can be used to inform our understanding of the characteristics of texts in different genre, this is called genre analysis and was devised by Swales (1990, 2004). Drawing on corpus-based methods, many researchers have used the move analysis framework to analyse sections of research articles or theses, such as previous studies by Samraj (2013, 2016), Hyland (2004), and Kanoksilapatham (2012).

On the micro-level of linguistic analysis, analysis of collocations derived from Firth's collocation concept (1957:11) "You shall know a word by the company it keeps" has emerged. The concept of collocations or collocability refers to the investigation of individual words and the words that surround them (usually a span of 1 to 5 words).

Based on the collocation concept, phraseological multi-word sequences, or lexical bundles, have been investigated using different analyses (see examples in Chapter 3). Lexical bundles is one of the umbrella terms within phraseology (Biber & Conrad, 1999; Biber, 2006; Biber & Barbieri, 2007). It is based on extended collocations or the study of continuous 3–6-word bundles based on frequency and dispersion range (Biber, Conrad, & Cortes, 2003, 2004; Cortes, 2015). The lexical bundles framework has been used widely to study academic corpora, either investigating within a single corpus or comparing within or among corpora (Nesi, 2016). Moreover, lexical bundles and corpus techniques can be used in classroom teaching, as seen in Lewis (2000), Cortes (2006), Kazemi, Katiraei, & Rasekh (2014). Despite previous corpus studies research, it is challenging to identify the potential use of corpora for teaching English in a specific field like engineering (Biber & Gray, 2010). This highlights the importance of examining the context and theoretical foundations of this study, as well as the structural and functional approaches used to explore how corpus-based methodologies can enhance engineering-specific English instruction. The following sub-section discusses the significance of this study and its alignment with existing research while outlining the theoretical positioning and structure of the thesis.

1.1.1 The context and setting

As mentioned earlier, the primary aim of this research is to help develop ways to assist students in writing research articles in the engineering field. As aforementioned, corpus pedagogy studies have been applied to students doing English for Academic Purposes (EAP), such as identifying the creating a research space (CARS) model by Swales (1990, 2004) and lexical bundles' use in research articles moves analysis by Cortes (2013). The CARS model provides a structure for organising the introduction of academic papers. It consists of three main moves: (1) establishing a territory, (2) establishing a niche, and (3) occupying the niche. These moves guide authors in clearly positioning their research within the existing literature by setting the context, identifying gaps, and presenting the research contribution (Swales, 1990, 2004). Most corpus pedagogy studies seem to fit well with students who are at an advanced or intermediate level of English proficiency, such as master's or PhD students (cf. Lee & Swales, 2006). Undoubtedly, these groups of students seem to be able to acquire knowledge quickly, since they have a strong background in English skills, as seen in studies by Charles (2014, 2015).

In this context, the subjects in this research range from lower-intermediate to intermediate levels of proficiency.[1] In relation to Kasetsart University, Chalermphrakiat Sakon Nakhon Province Campus (henceforth KU.CSC) English language requirements, the subjects are 3rd-and 4th-year engineering students who are required to write research articles for their engineering projects before graduation. However, the students' ability to use English in their writing might affect their success. I witnessed this situation in 2011, where many students struggled with writing research articles due to limited English proficiency, particularly in using technical terms and constructing coherent academic arguments. Therefore, I decided to develop a possible corpus pedagogical approach that would be suitable to address students' needs when writing research articles and to develop a sample of materials that are drawn from the lexical bundles' structure and function used

[1] Specifically, in this context, the level of English proficiency of students in Thailand may not be comparable to students in the US or the UK, so that the English ability specified here might be lower.

in the ERAs. More importantly, this approach can be used to develop and be embedded in the current syllabus for English for Specific Purposes (henceforth, ESP) and teaching engineering students at KU. CSC, particularly regarding writing research articles.

Moreover, my intrinsic motivation came from my experience of teaching engineering students at KU. CSC. In 2011, one of the most critical situations that happened was when a group of engineering students dropped into my office and said, "Could you please edit the abstract of the article for me?" When I looked through it, most of the words were structurally combined into a subject-verb-object pattern. However, they were all meaningless fragments and sentences and comprised some erratic uses of word-for-word translations. At that time, I thought that two things might be going on with engineering students writing research articles.[2] First, perhaps the students know the main structures of English (these appear to be subjects, verbs and objects in sentences). Possibly, the students used knowledge gleaned from the English courses[3] provided. Even though they had done some English courses, they could not successfully convey meaningful messages. Based on the students' written sentences, I wondered why they could not produce meaningful or natural sentences, even though they had taken English courses, such as Foundation English[4] and English for Specific Purposes (ESP).[5] Some of the sentences produced by the students confused me, and I assumed that they relied on translation from their first language into English. Also, the fragments or sentences produced relied on word-for-word translation and may have been translated using the Google Translation Tool. In terms of the English coursebooks provided, the ones currently used in teaching English Foundations are New Interchange (Richards et al., 1997) and Face2face (Cunningham et al., 2013). The

[2] The engineering students are required to write research articles in one of their compulsory courses, run as seminars. The seminars involve reading research articles relevant to engineering projects in specified journals, presenting papers in the classroom, and writing parts of research articles at the end of the engineering course at Kasetsart University (KU) (see also English courses for engineering students in Appendix A)

[3] One of the compulsory English courses provided, IV, is called 'English Structure'. It is usually for students who have passed English Foundation I, II and III.

[4] English compulsory courses, i.e.,Foundation English I, II and III, are for undergraduate students in all majors and aim to teach essential English communication,

[5] English compulsory courses for undergraduate students who pass English Foundation I, II and III. ESP courses provided at KU. CSC include English for Business, Technical English and English for Communication.

ESP course instructors are responsible for producing learning materials that match the course syllabi.

Regarding Foundation English I, II and III, the primary purpose of using commercial coursebooks is to enable learners to communicate in English. Moreover, the use of coursebooks might decrease the workload of the teachers, in the hope that they will give students a better sense of English (Dat, 2008; Masuhara, 2010; Richards, 2015). The coursebooks contain the linguistic features and functions of English used in school, work, social life and leisure situations. There are four key English skills: writing, reading, listening and speaking. Vocabulary items are presented as words, parts of speech and meanings in situations to prepare the context of lessons for learners. Depending on the lesson situation, grammar points are used to indicate actions and time in the lesson situation, e.g., the present simple is used to describe daily activities in school, work or hobbies. Listening strategies, like listening for specific information and listening for main ideas, are taught. Regarding reading skills, short passages and reading strategies such as skimming and scanning techniques can also be observed in coursebooks' activities. Concerning speaking skills, students can give opinions on specified situations by using the functions and patterns provided. The coursebooks are visually appealing and up-to-date in design, with many colourful pictures that engage students and make the material more accessible. Regarding culture, they offer four English skills in the Western context that the students can learn from (Tomlinson & Masuhara, 2010; Tomlinson, 2011).

After the students pass all the Foundation English courses, they can choose to study English for Specific Purposes (IV). ESP courses focus on English aspects that are used in specific contexts, such as technical, professional, or academic texts. The main goal of these courses is to enable students to read, listen to and write academic texts. Moreover, the teachers have to prepare subject-matter content related to engineering for this subject based on the course syllabus provided. Currently, a few courses are provided for undergraduate students (3rd and 4th years), and I would like to mention one of the ESP courses called 'Technical Writing', which has been used to teach writing to engineering students (see Appendix A). Based on my analysis of the 'Technical Writing' course in January 2016, the

materials that are currently used in the 'technical writing' course support reading academic texts and doing grammatical error cloze tests. Short reading passages are not very relevant to the academic texts that the engineering students use in their learning. Also, the texts used in the materials seem to be below most students' level of proficiency. And the students can select answers without needing any complicated thinking; they can easily guess the answers. The grammatical points covered in the sessions do not match the contexts of lessons. Instead, they test the students on rules of grammar (e.g., 12 tenses in English) without considering the context.

There are advantages in using commercial coursebooks, such as covering the four main English skills, learning English from Western cultures and seeing language features and functions used in context. However, to some extent, the students might not have adequate knowledge of using language in real situations. For example, in the vocabulary sections, the students are taught individual words with limited examples. Just some words and their meanings used in context are presented without giving many samples of how words are used. Moreover, the words and grammar that occur with target words in context are not usually provided in commercial coursebooks. Again, the best way to learn words is by learning from the words that surround them, too (cf. Firth, 1957: 11). The cultural aspects covered in the coursebooks might not be familiar to the students, since they have never lived abroad. Regarding culture, the students might find it hard to connect their (Thai) culture with the one used in the coursebooks (e.g., the US or the UK), Masuhara (2011: 236) refers to this as ‘alienation of the coursebooks’.

Similarly, there is no in-depth focus on specific skills and contexts. To give an example from the vocabulary section, the teaching of individual words provided in the technical writing course seems to involve providing words and meaning patterns without samples of the words in use. The vocabulary activities do not require much thinking skill to choose the right words in context. In the same way, the students can find the answers in a dictionary. Grammatical error identification activities might not prepare the students to write ERAs as required from the engineering department afterwards. Hence, the use of commercial coursebooks and in-house materials might not help to prepare students for

writing research articles in engineering (cf. Chen, 2011). To help support these students, there is a need to make changes to the materials provided and the process of developing the programme to suit the students' needs, proficiencies and styles.

1.2. Significance of the study, objectives and research questions

Writing is a complicated process and requires knowledge of vocabulary, grammar and discourse (Alderson, Clapham, & Wall, 1995). Writing in scientific contexts requires more than a general knowledge of lexis and grammar (Hyland, 2004). Instead, some concepts that are used to explain the scientific concepts are composed of certain words and expressions (cf. Strevens, 1973). Teaching learners who are in a subject-specific discipline to write research articles is challenging because many factors need to be taken into account such as the corpora to be used and methods in delivering the teaching (cf. Charles, 2015). One of the reasons for choosing engineering students is pointed out by Anthony as below:

> [Teachers in] Asian countries are most comfortable using a teacher-centred, prescriptive approach to writing instruction, in which they give students a set of rules (or guidelines) to follow, for example, in terms of paragraph structuring, choice of logical connections or *phraseology*. Although this may be partially successful in a general English writing classroom, its success will be very limited in a *discipline-specific writing classroom* due to the heterogeneous group of learners […] it is still questionable if a teacher-centred, prescriptive methodology should be utilised.[6]

(see Anthony, 2017: 162; emphasis added)

The above quote implies that to be able to teach writing to discipline-specific students, there is a need to consider the methodology for teaching this group of students. Writing research articles requires knowledge of the subject to be written about, linguistic

[6] In this study, to distinguish long direct quotes from other ones, they are indented. Within each quote, I use italics to emphasise or support the most important points for this research.

knowledge, i.e., syntactic and discourse functions, and ways in which to present the content, i.e., rhetorical move knowledge that can be recognised in communities (Hyland, 2000; Lillis & Curry, 2010). Another reason for choosing lexical bundle analysis is given by Gray below:

> [Thus] *lexical bundles* are seen as basic components of discourse construction which can help language users carry out particular functions.
>
> (see Gray, 2016: 33; emphasis added)

Moreover, one of the reasons I choose lexical bundles is because it is useful for students to learn English in chunks, and there are units of meanings in every single one (Pawley & Syder, 1983). The use of lexical bundles is beneficial for students and accessible for them when acquiring a language, especially for second language learners (Granger & Paquot, 1998). Corpus-based methods play a crucial role in identifying and classifying lexical bundles.

The lexical bundles taxonomy was pioneered by Biber and others. They have developed lexical bundles and extended collocations from the concept of Firth's (1957) collocation study. The lexical bundles framework has been used extensively in analysing academic corpora, such as research articles and theses sections by Cortes (2013) and Hyland (2008), in spoken and written registers by Douglas Biber and Barbieri (2007), Douglas Biber, Conrad, and Cortes (2004) and Csomay (2013), in specific disciplines such as history by Bal (2010) and Cortes (2008).

This study focuses on the analysis of lexical bundles that are used in engineering research articles (see examples in Chapter 3). The analysis of lexical bundles' structure and discourse functions is used to inform a sample of materials used to teach engineering students (see examples in Chapter 4). A corpus approach, data-driven learning, is used as a tool to train students to study target lexical bundles' structure and functions (see examples in Chapter 4). DDL involves using corpus linguistics tools to allow learners to engage directly with authentic language data, which helps them observe real-world language patterns and improve their understanding of complex structures (Boulton, 2017). This study

also addresses the process of development of materials as well as students' attitudes towards the use of corpus techniques (e.g., concordances, AntConc 3.2.4 from Anthony, 2011) to study target bundles (see examples in Chapter 4). The prime objective is to answer the following guiding question:

> *In what ways can engineering students learn to write research articles by using lexical bundles and materials that are related to lexical bundles?*

Thus, the research questions derive from the broad scenario described above:

> **RQ 1)** What are the generic types of lexical bundles (Top 50) and specific lexical bundles that are more relevant to phrasal expression in the engineering domain (Bottom 50), as used in engineering research articles? (Chapter 3)
>
> **RQ 2)** How are these generic and specific lexical bundles used in engineering research articles, regarding frequency, dispersion range, structure and discourse functions? (Chapter 3)
>
> **RQ 3)** To what extent is using lexical bundles in a 'data-driven learning' approach effective for teaching English writing to engineering students in a Thai university? What are the strengths and weaknesses of using this approach with these students? (Chapter 4)

In brief, Research Questions 1 and 2 aim to explore the structure and discourse functions of lexical bundles and how they are actually used in engineering research articles. Here, *structure* refers to their grammatical form, *function* refers to their communicative role in the text, and *use* refers to how these bundles are employed in real examples across different contexts. RQ 3 aims to inform a sample of materials for teaching engineering students and how to use the materials to teach engineering students drawing on corpus methods, called data-driven learning. To address the first two research questions, I created a corpus from 100 engineering research articles that have been used for the teaching and learning of

engineering students at KU.CSC, see methods of building a corpus from Henry & Roseberry (2001), Hunston (2002), McEnery et al. (2006), and Baker (2006, 2010).

I conducted a quantitative analysis of lexical bundles to identify and classify their structure and functions using the lexical bundles taxonomy framework from Biber et al. (1999) and Conrad & Biber (2005). Similarly, I was able to identify lexical bundles that are in frequent or general use in an engineering research article corpus. To be able to study lexical bundles that are specific to the engineering discipline, I initially set minimum thresholds of frequency ≥ 2 and dispersion ≥ 2 texts, then increased these thresholds step by step until the list was narrowed to exactly 50 3-word and 4-word lexical bundles (see examples in Chapter 3).

To answer my third research question, first, I focused on the target lexical bundles to be taught in the assembled materials and developed a sample of materials. The content of the materials is shaped by the structures and functions of the lexical bundles. Trialling and revising the materials was also conducted at this stage (see examples in Chapter 4). Instruction in lexical bundles, with the aid of corpus methodology, was provided. The results from teaching based on students' pre-test and post-test scores, students' attitudes towards lexical bundles and the materials used in teaching are used to answer RQ.

1.3. Theoretical positioning

A considerable amount of literature pays attention to lexical bundles, as can be seen in Csomay (2004), Nesi and Basturkmen (2006), Biber and Gray (2010), Ädel and Erman (2012) and Gray (2016), see examples in Chapter 2. Studies on lexical bundles are situated within research on phraseology. Phraseology was established from two major ideas: the British tradition (Firth, 1957 & Sinclair, 1991) and the traditional Russian approach (Altenberg, 1998 & Cowie, 1998). While the British tradition has been interested in words and meanings in context (Stubbs, 1993), the traditional Russian approach has been influenced by producing dictionaries based on lexicon and semantic methods, known as word-like and sentence-like, as can be seen in Mel'čuk (1998). In Mel'čuk's approach, the term *word-like* refers to lexical units that resemble single words, while *sentence-like* refers

to larger multi-word expressions that resemble sentences in their syntactic structure. This distinction can be seen in the works of Mel'čuk (1998), who developed a theoretical framework for lexical functions, emphasising the relationship between words and their semantic roles within these categories. These aspects reflect a structured analysis of language at both word and sentence levels. All these traditional studies mentioned are based on the frequency-based approach, and it is the central interest of this current research.

1.3.1. Definition and concept of lexical bundles

Lexical bundles are contiguous sequences of three, four, five or six words that appear together in a fixed order, like *in terms of* or *on the other hand* (Cortes, 2004). Lexical bundles are also different because they often bridge two clauses or phrases rather than forming a complete structural unit on their own. In this study, I follow the concept of lexical bundles as defined by Biber et al. (1999), who set a frequency threshold to identify bundles. This means that for a sequence of words to be considered a *lexical bundle*, it must appear frequently in a corpus. Other terms like *cluster* or *n-gram* do not require such a frequency threshold to be defined.

For example, in academic writing, common lexical bundles include phrases like *in the context of* or *as a result of*. In contrast, conversational English might include bundles like *you know what I mean* or *at the end of the day*. These bundles help structure language in specific contexts and are useful for language learners to recognise. This study is mainly based on the concept of Biber et al.'s (1999) lexical bundles taxonomy framework. By analysing the frequency and patterns of these bundles, we aim to provide materials that help Thai engineering students understand and use lexical bundles effectively in their writing.

1.4. The structure and function of lexical bundles and their use in teaching

Here, I start by considering the structure and function of lexical bundles based on Biber et al.'s (1999) lexical bundles taxonomy. Lexical bundles are divided into 12 main structures. Since lexical bundles comprise phrasal and clausal fragments, I categorise the 12 structures into three main types, i.e., Noun-Phrase (NP) based, Preposition Phrase (PP) based, and Verb Phrase (VP) based (Biber et al., 2004). In terms of function, I use the classification

of lexical bundle functions from Biber et al. (1999): stance expression, discourse organisation and referential expressions (see examples in Chapter 3).

Several recent corpus-based and corpus-driven studies investigating phraseology have been carried out to identify lexical bundles' characteristics and discourse functions (cf. Csomay, 2013; Gray & Biber, 2015). Studies of lexical bundles' structure and functions have been conducted on a single corpus, between corpora and among corpora (Nesi, 2016). The use of lexical bundles has been examined in academic discourse, such as academic lectures by Nesi and Basturkmen (2006), academic spoken and written registers by Biber and Barbieri (2007), Biber and Conrad (1999a), Conrad (2015, 2017) and Gray (2016), academic papers such as master's theses and research articles by Hyland (2008) and specific disciplines by Cortes (2004).

I have adopted the use of lexical bundles with corpus techniques in teaching. There are studies that have applied lexical bundles' structure and functions to teach students in subject-specific disciplines, such as Cortes (2006) and Kazemi et al. (2014). In terms of materials production, the corpus techniques used in this study have been adapted from the concept of data-driven learning - see John (1991), Sripichan (2010), Anthony (2017).

1.5. Structure of the thesis

As mentioned earlier, the principal aim of this study is to examine the impact of using analyses of lexical bundles to support Thai engineering students learning to write research articles, as well as to develop a sample of materials for engineering students to use when writing research articles. To answer the overarching research question, a corpus from engineering research articles was built, and criteria for identifying lexical bundles have been used. Structures and functions were used to analyse and classify the top 50 3-word and 4-word bundles and the bottom 50 3-word and 4-word bundles. Then, materials informed by lexical bundles' structures and functions were developed and used with engineering students. To make the lexical bundles' categorisation valid, the inter-rater reliability used in the lexical bundles identification of structures and functions was adopted from Gray (2016). Thus, this thesis consists of five chapters. The first chapter outlines the

significance of this study, the reasons for choosing lexical bundles and engineering students, the research questions, and the methodology for analysing lexical bundles, in particular the concepts of the lexical bundles taxonomy framework from Biber et al. (1999).

Chapter Two of the thesis starts with a brief account of corpus linguistics, paying special attention to the use of corpus techniques in analysing academic corpora. The chapter also considers the concepts of phraseology that lexical bundle methods are based on, i.e., collocations and a frequency-based approach. Moreover, Chapter Two describes the lexical bundles concept, methods and relevant research findings. Given that lexical bundles are the main approach to the present study, this chapter describes the concept, methods, structures and discourse functions of lexical bundles. The chapter also includes the use of lexical bundles' structures and functions in language learning and teaching.

Chapter Three presents the data and the methodological framework. It adopts Biber et al. (2004) and Biber et al. (1999) as its core analytical framework. The chapter explains the reasons for the criteria used in lexical bundle selection. The structures and functions of lexical bundles are drawn from Biber et al. (2004) and Biber et al. (1999). The use of inter-rater reliability in lexical bundle analysis is adapted from Gray (2016), and Mackey & Gass (2015).

Regarding structure, 12 main structures from Biber et al. (1999) are used to categorise the top 3-word and 4-word bundles and the bottom 50 3-word and 4-word bundles, 200 in total. These 200 lexical bundles are grouped into three main structural types: noun phrase (NP)-based, preposition phrase (PP)-based and verb phrase (VP)-based (see Biber et al., 2004). In terms of function, stance expression, discourse organisation and referential expression are used as the main lexical bundle function classification.

The main aim of Chapter Four is to answer RQ3. The present study is interested in using lexical bundles' structure and functions in teaching writing research articles to Thai engineering students. This study employs data-driven learning and other corpus techniques (e.g., concordance), aided by a facilitator and paper-based materials. In the same way, Chapter Four explains the process of development of research article writing materials and allocating teaching time to lexical bundles. It also includes the methods in data collection,

such as ethical approval, pre-test, post-test, using data-driven learning in teaching and research instruments. Inter-rater reliability is addressed concerning analysing students' use of lexical bundles in pre-tests and post-tests. The results from teaching lexical bundles and the use of lexical bundles in students' pre-tests and post-tests are discussed, including students' attitudes towards learning lexical bundles provided in a workshop.

Finally, in the concluding chapter, the main findings and contributions are summarised, and this is followed by a discussion of the limitations of the study (§5.6), suggestions for directions for further research on lexical bundles, the application of lexical bundles in teaching (§5.7) and concluding remarks (§ 5.8).

2 Chapter Two: Corpus linguistics, phraseology, lexical bundles and application of lexical bundles in ESP instruction

2.1 Introduction

Corpus linguistics offers a systematic way to analyse language features and patterns which characterise samples of particular kinds of language (cf. Hunston, 2002; Orr, 2006; McEnery & Gabrielatos, 2008; Leech, 2015). The results of corpus analysis do not only prove to be useful resources for linguists, researchers and practitioners; they can also contribute to language teaching (cf. McEnery, Baker, & Wilson, 1995; Leech, 1997; Braun, Kohn, & Mukherjee, 2006; Conrad, 2015, 2017). This chapter cannot describe in detail the growing body of research into corpus linguistics, as it is too large to fit within the limits of this study. The aim, instead, is to give a critical account of corpus-based research that is most closely related to the present investigation. Thus, as a first step, this chapter provides a definition and describes the concepts and use of corpus techniques in analysing academic corpora (§ 2.2). Based on the Russian tradition and the Neo-Firthian tradition, the sub-sections of 2.3 include an account of the methods and criteria for identification of recurrent word combinations (§ 2.3.1). Previous research on phraseology is presented in 2.3.1.3. This chapter further considers lexical bundles as one way of analysing phraseological units, which is a central focus of this current study (§ 2.4). Drawing on the work of Biber, Finegan, Johansson, Conrad, and Leech (1999) on lexical bundles' structure and discourse function, the chapter further outlines structures, discourse functions, and empirical studies on lexical bundles that are used as the main framework in this current research. Section 2.5 explores research findings on the application of language corpora and phraseology in language pedagogy. Finally, Section 2.6 concludes the chapter by summarising the research.

2.2 Corpus linguistics: definition, concepts and use of corpus techniques in analysing academic corpora

To understand the term 'corpus linguistics', it makes sense to discuss the definition and concepts of corpus linguistics debated by scholars. I also examine the corpus techniques

used to understand academic discourse for the sake of this research. In the digital age, corpora and corpus software have become powerful resources used by researchers, linguists and practitioners for their ability to store and process language samples (cf. McEnery & Wilson, 1996; McEnery & Gabrielatos, 2008; Prodromou, 2008). As its Latin root suggests, the term 'corpus' refers to a "body of texts" (Baker, 2010; McEnery & Wilson, 1996). There is general agreement on the meaning of the term 'corpus' as a 'collection of naturally occurring language text, chosen to characterise a state or variety of a language' (Sinclair, 1991: 171; see also [Baker, 2010; Biber, Conrad, & Reppen, 1998b; Flowerdew, 2010; Flowerdew, 2012; Hunston, 2002; McEnery & Hardie, 2011; McEnery & Wilson, 1996, 2001; Mukherjee, 2006; Nesi, 2016; Prodromou, 2008; Sinclair, 1991; Teubert & Cermáková, 2007]). Regarding this methodological approach, researchers can investigate different perspectives on language variation and use (Biber & Reppen, 2015; McEnery & Gabrielatos, 2006). Quantitative methods involve the use of statistics in identifying the frequency of occurrence and co-occurrence of language items (McEnery & Gabrielatos, 2008). Qualitative methods involve closer interpretation of data, e.g.,using techniques drawn from discourse analysis (cf. Baker, 2006).

Two approaches are usually debated in corpus linguistics: corpus-based and corpus-driven approaches (McEnery & Gabrielatos, 2008; Biber & Rippen, 2015). To distinguish between these two terms, Tognini-Bonelli (2001: 65) explains the term *corpus-based approach* as follows:

> …the corpus is mainly used to "expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study".

As can be seen from the above definition, the techniques used in this approach can lead to dismissing data that do not fit with the theory or making data fit the theory (Biber & Reppen, 2015; McEnery & Gabrielatos, 2008; McEnery, Xiao, & Tono, 2006). By using data annotation as the main technique, researchers can reduce data to "a set of orderly categories which are tractable within existing descriptive systems" (Tognini-Bonelli, 2001:

68). When using such a technique, the influence of preconceived theory and the annotator's intuition have been criticised (McEnery & Gabrielatos, 2008; Tognini-Bonelli, 2001).

In contrast, a *corpus-driven approach* aims to identify 'facts about language free from the influence of existing theoretical frameworks, which are considered to be based on intuitions' (McEnery & Gabrielatos, 2008:36). Furthermore, the main goal of a corpus-driven approach is to identify 'the patterning of orthographic words' from a corpus (McEnery & Gabrielatos, 2008: 36). To sum up, a corpus-based approach aims to find evidence (from data) for or against a given theory, while a corpus-driven approach aims to observe patterns in corpus data in order to derive insights into language without relying on pre-existing theories and frameworks (McEnery & Gabrielatos, 2008).

Apart from methods for analysing text, machine-readable corpora have been developed for a wide range of research purposes. Currently, many corpora have been created to serve key objectives in studying languages. First, corpora have been designed to study the languages used by national or regional populations, such as the American National Corpus and the New Zealand Spoken English Database (Orr, 2006). One well-known corpus is the British National Corpus (BNC) – 100,106,008 words in size – which is designed to represent a wide range of British English (McEnery et al., 2006). In addition, there are corpora that have been created to study specific kinds of language use in particular professions, such as the Hong Kong University of Science Computer Science Corpus (cf. Jame, 1993), the Guangzhou Petroleum English Corpus (cf. Qi-bo, 1989), the Student Engineering English Corpus, the Corpus of Professional English and the Longman Learners' Corpus.

Researchers can also build corpora themselves, depending on the purposes of their research. This may be due to the unavailability of suitable corpora. Do-it-yourself (DIY) corpora has become one of the choices for researchers. By using appropriate software (e.g.,WordSmith Tools, AntConc), statistical information can be extracted from compiled corpora. The statistics derived (e.g., frequency, dispersion range, keyness, log-likelihood) and patterns of concordances can be used as the basis to identify the characteristics of language use in the target genre (Hyland, 2015; Tognini-Bonelli, 2001). There are some

guidelines that the researcher should be aware of, such as considering the purposes of creating a corpus, appropriate corpus size and the representativeness of a corpus (Mukherjee, 2006). There are no fixed rules about the size of a corpus, because a large corpus may not be essential as long as the size is appropriate for the purposes of the research (Baker, 2010). A small corpus may have sufficient examples of frequent linguistic features. Biber (1994) notes that, in some cases, a sample of 1,000 words may prove adequate. Again, the research purpose plays a crucial role in choosing methods from corpus linguistics (Hunston, 2002).

2.2.1 The use of corpus techniques in analysing academic corpora

Many studies have used corpus techniques to study a wide range of academic corpora via qualitative and quantitative approaches. By using two approaches, corpus linguistics has contributed to the understanding of language in academic discourse from *lexical, grammar and multi-word combination perspectives* (McEnery & Gabrielatos, 2008; Nesi, 2016). Regarding quantitative methods, researchers often use frequency to reveal the regularities of selected data. The investigation of meaning and communicative function is a way to make qualitative judgements about the target dataset (Hyland, 2015; Nesi, 2015; Tognini-Bonelli, 2001).

It is important to acknowledge that it is difficult for a study to be either fully corpus-based or fully corpus-driven. McEnery and Hardie (2011) highlight the complexity of these methodologies, suggesting that they frequently overlap. The validation of hypotheses or linguistic theories is typically achieved through corpus-based approaches, which rely on pre-existing frameworks (McEnery & Hardie, 2011). They are widely used in the fields of lexicography and grammar studies, where they offer quantitative data and authentic examples to enhance teaching materials. On the contrary, corpus-driven approaches focus on discovering linguistic patterns directly from corpus data, emphasising inductive analysis and exploratory learning. For instance, data-driven learning (DDL) promotes critical thinking and autonomous learning by encouraging learners to actively interact with corpus data. The objective of this study is to identify linguistic patterns in engineering research articles (ERAs) and apply these insights to inform teaching practices, which is more closely

aligned with a corpus-driven approach. Nevertheless, it also integrates corpus-based elements by using pre-existing linguistic frameworks to analyse the data.

Vocabulary or individual words have been considered important in academic writing as they contain both meanings and grammar in context (cf. Wilkins, 1972; Coxhead, 2015). In the investigation of individual words or vocabulary, frequency-count techniques have been used to identify frequent lexical items. For lexicographical purposes, a dictionary is first built based on frequent lexical items. For example, the BOE [the Bank of English] created by Birmingham University in the 1980s was used to provide information for the Collins Cobuild English Language Dictionary (CCELD) (Moon, 2009; Nesi, 2015; Sinclair, 1987). The CCELD, first published in 1987, serves as a vast resource for vocabulary learning, as Nation (2013) notes.

A comparison of frequency counts within or between corpora can be used to identify a 'simple word from more complex ones' (Nesi, 2016: 211). Academic wordlist research has been conducted to support university students in their learning, mainly in the reading of textbooks, articles, book chapters and laboratory work. A well-known academic word list that has applied the frequency method is the Academic Wordlist (AWL). The Academic Word List (AWL), created by Averil Coxhead in 2000, is an essential resource for English as a Second Language (ESL) learners. The AWL comprises 570-word families that often occur in academic writings, making it particularly helpful for students aiming to succeed in an English-speaking academic context. In contrast to West's General Service List (GSL, West 1953), which comprises high-frequency terms for general application, the Academic Word List (AWL) is primarily focused on academic vocabulary. The AWL provides several significant benefits for learners. Initially, it contains words that are commonly used in scholarly literature, representing around 10% of academic vocabulary. Secondly, the AWL is varied, including fields such as Arts, Commerce, Law, and Sciences. Third, it is divided into 10 sublists according to word frequency, enabling learners to enhance their vocabulary gradually. Ultimately, it allows both independent study and organised instruction, making it an important resource for educators and learners in English for Academic Purposes (EAP) programs. Proficiency in the AWL enhances understanding of

academic literature and markedly boosts performance in academic reading and writing activities.

The AWL has been applied to study language in discipline-specific areas to help learners overcome difficulties in studying language in a specific domain, and to increase awareness of the use of subject-specific vocabulary in the field (Coxhead, 2011, 2015; Nesi, 2015). Recent work derived from the AWL methodology can be seen in the work of Martínez, Beck, and Panza's (2009) academic vocabulary in agriculture research articles, and Coxhead and Hirsch's (2007) Science Word List for EAP. Other studies on academic word lists have been conducted in specialised fields such as engineering (Mudraya, 2006; Ward, 1999, 2009) and medicine (Hsu, 2013; Lei & Liu, 2016; Wang, Liang, & Ge, 2008).

Based on the corpus techniques used in deriving Academic Wordlists (AW), some studies have integrated other corpus techniques to identify *core academic vocabulary*, as seen in Paquot's (20110) Academic Keyword List (AKL). AKL contains 930 items that are frequent in a corpus of academic writing when compared to a reference corpus. Similarly, Simpson-Vlach and Ellis (2010) generated formulaic sequences (lexical bundles) from academic speech corpora (e.g., MICASE and BNC) and academic writing corpora (from a selection of research articles in Hyland's (2004) research-article corpus and BNC files). Simpson-Vlach and Ellis (2010) used both qualitative and quantitative techniques. Quantitative methods included the use of statistical measures (frequency, mutual information) to identify lexical bundles based on studies by Biber and Conrad (1999) and Biber, Conrad, and Cortes (2004). The use of judgements by experienced EAP instructors and language testers can confirm whether target lexical bundles are meaningful.

For EAP study purposes, corpus techniques seem to be promising for studies of *discipline-specific language* as they can investigate the language used in a specific field, such as, technical terms, meanings and collocations (Hyland & Tse, 2007). Nesi (2015: 212), in her work on corpus studies in EAP, notes that words behave differently in different disciplinary domains as regards meanings and collocations. Lexical bundles have been compared with subject-specific corpora by many researchers using corpus techniques. Previous research comparing lexical bundles across fields is found in Cortes (2004), on

history and biology, in Hyland (2008), on applied linguistics, biology, business and electrical engineering, and in Salazar (2011), on scientific writing.

Studies by Leech (2015) and Hunston (2015) show the importance of grammar and lexis, which are inseparable in language. In particular, these studies highlight how grammar provides structure to language, while lexis conveys meaning, and their interdependence is crucial for understanding linguistic patterns. For example, Hunston (2015) explores the role of lexico-grammatical patterns in conveying academic discourse, while Leech (2015) examines the interaction between lexical choice and grammatical structures in both written and spoken contexts.

A review of words and grammar in four major registers (fiction, conversation, new reportage, academic prose) can be seen in Longman's Grammar of Spoken and Written English (LGSWE) [7](Biber et al., 1999). The LGSWE is based on a model of English grammar constructed in 1972, and extended in 1985; it was developed by Quirk et al. (1972) with reference to a 40-million-word corpus. By using the frequencies of occurrence of words, LGSWE describes the use of phrasal words, tenses and modal verbs, and their frequencies. It also allows comparisons between verb-phrase types in each register and between registers (Cortes, 2015; Hunston, 2015). In terms of quantitative information, LGSWE represents the grammatical features of lexical bundles and how they are used in conversation (such as, *I don't know why*, *I thought that was*) and academic prose (such as *the nature of the*, *as a result of*), with registers based on frequency of occurrence (Biber et al., 1999).

So far, this section has presented a definition and concepts of corpus linguistics, as well as the use of corpus techniques in analysing academic corpora for English for Academic Purposes (EAP). It briefly discusses relevant studies that have applied corpus techniques qualitatively and quantitatively to study language varying from individual words to complex structures, such as lexical bundles, multi-word combinations and grammar. While

[7] In this study, the description of lexical bundles' structure and functions in academic prose is mainly based on the LGSWE (Biber, Finegan, Johansson, Conrad, & Leech, 1999). The term 'lexical bundles' is based on the terms, concepts and criteria (e.g.,frequency and range dispersion) in Biber, Conrad, and Cortes (2003); Biber et al. (2004); Biber et al. (1999), (§ 2.4).

this study predominantly employs a corpus-driven approach to identify lexical bundles in Engineering Research Articles (ERAs), it also integrates corpus-based elements. Specifically, the use of pre-existing categories such as prepositional phrases (PP) and noun phrases (NP), in conjunction with functional classifications, represents the implementation of corpus-based techniques. But overall, this integrated approach situates the research within the corpus-driven spectrum, where inductive pattern identification is enhanced by established linguistic frameworks to ensure both uniqueness and alignment with linguistic theory. This section also explains corpus techniques that enable comparisons within or between corpora to study how words behave differently in different texts (cf. Nesi, 2016). In the course of this study, I will focus in particular on frequency and distribution criteria to guide the selection of lexical bundles in ERAs (see examples in Chapter 3). The structure and functions of lexical bundles are also discussed along with associated frequencies (see examples in Chapter 3).

While corpus linguistics offers methods to identify linguistic features quantitatively and in different registers and can observe the language use from instances in concordance lines, there should be approaches/ frameworks to identify and categorise derived linguistic features (cf. Hunston, 2002). The current study employs corpus linguistics to identify linguistic features (i.e., lexical bundles) and this approach is used in lexical bundles classification. Since the concept and methodology of lexical bundles identification are situated within a wide spectrum of phraseological studies (cf. Howarth, 1996; Wray, 2002; Gray & Biber, 2015), it makes sense to discuss relevant approaches, concepts and how phraseological units are classified (§ 2.3). In particular, Section 2.3.1 briefly describes two theoretical perspectives that are relevant to aspects of identifying phraseological units: the Russian tradition (cf. Arnold, 1986; Howarth, 1996; Cowie, 1998) and the Firthian tradition (cf. Sinclair, 1991; Howarth, 1996). The two approaches assist in a methodology design for the identification of lexical bundles' form and function in an ERAs corpus (for a more detailed discussion, see sections 2.3.1.1and 2.3.1.2). Research findings based on the two phraseological approaches are discussed in Section 2.3.

2.3 Phraseology: concepts, methods and previous research

In everyday use, language is full of 'prefabricated expressions rather than being strictly compositional' (Gray & Biber, 2015: 125), it is composed of a variety of units, including syntactic units (e.g., phrase, clause, sentence), language functions (e.g., greeting, farewell) and proverbs (e.g., When in Rome, do as the Romans do) (cf. Howarth, 1996; Granger & Paquot, 2008). These units of language are combined and called word combinations (cf. Howarth, 1996; Cowie, 1998). Moreover, Altenberg (1998) in his investigation of the London-Lund Corpus (cf. Svartvik, 1990), found that 80 per cent of words in the corpus formed part of recurrent word combinations.

Collostructions are important in phraseology because they show how words and grammar patterns work together (Gledhill, 2011). The term comes from construction grammar and refers to the strong connection between certain words and specific grammatical structures. For example, some words are more likely to appear in passive constructions, while others prefer active forms. These patterns are based on frequency — the more often a word appears in a certain structure, the stronger the collostruction. In both native and learner writing, collostructions help us understand how language is used, especially when it comes to idiomaticity and natural expression. Research shows that learners sometimes use uncommon or awkward collostructions due to influence from their first language or a lack of awareness of register. This highlights the importance of studying collostructions to help students develop better phraseological competence (e.g., Garner, 2018; Gilquin, 2012).

The study of phraseology may refer to the nature of word combinations, i.e., form, and how these structures are used to convey communicative purposes, i.e., function (cf. Howarth, 1996; Wray, 2002; Cortes, 2015). Thus, various terms have been used to refer to different forms and methods of phraseological classification: recurrent word combinations (cf. Howarth, 1996), formulas (cf. Granger & Meunier, 2007), multi-word expressions (cf. Moon, 1998), formulaic expressions (cf. Simpson-Vlach & Ellis, 2010), formulaic sequences (cf. Schmitt & Carter, 2004), lexical phrases, set phrases, fixed phrases, word-combinations (cf. Mel'čuk, 1998), fixed expressions (cf. Svensson, 2008), phrasal lexemes

(cf. Moon, 1998), prefabricated patterns, prefabs or lexical phrases (cf. Bolinger, 1976; Nattinger & DeCarrico, 1992; Granger, 1998), n-grams (cf. Stubbs, 2001), clusters (cf. Hyland, 2008), prefabricated patterns (cf. Hakuta, 1974), sentence stems or speech formulas (cf. Pawley & Syder, 1983) and lexical bundles (cf. Biber et al., 1999; Cortes, 2002; Biber & Barbieri, 2007).

As previously mentioned, phraseology examines combinations of various units (e.g., composite units, pre-patterned expressions). There is a blurred distinction between lexicon and grammar (Altenberg, 1998: 101). Thus, it is difficult to delimit these areas and classify different types of phraseology. In the same vein, Altenberg (1998: 110) states that lexicalisation and productivity are a matter of degree rather than a clear-cut division. This idea relates to Halliday's view of lexicogrammar, which sees lexis and grammar as working together to create meaning. Halliday (1994) explains that lexicogrammar reflects how word and grammar choices are connected, showing that language is used to express meaning rather than just follow rules. As can be inferred from this definition, lexicon and grammar are parts of formulaic sequences, and they normally overlap. Formulaic sequences can have *variability* (Wray, 2002). Variability in formulaic sequences refers to how these expressions adapt in form and function based on context, speaker intent, or communicative needs. For instance, a variable formulaic sequence is the phrase *at the end of the day,* which may appear as *by the end of the day* or *toward the end of the day* depending on contextual usage. This adaptability highlights the dynamic nature of formulaic sequences in real-world language use.

Therefore, different methods are used in identifying phraseology: corpus methods, traditional approaches and methods for the classification of phraseological units, which will be clarified in Section 2.3.1.

Regarding the corpus methods used in phraseology, as discussed above, two corpus approaches have been used to identify language variation and use: corpus-driven approach and corpus-based studies (§ 2.2 for a discussion of corpus linguistics concepts and methods). In phraseological identification and analysis, two methods are normally employed (Flowerdew, 2012; Gray & Biber, 2015). The first approach concerns the

researcher's pre-selection of phrasal expressions that s/he is 'perceptually salient or theoretically' interested in. They then further analyse the target phrasal expressions' use in the corpus (cf. Baker et al., 2006; Gray & Biber, 2015: 126). On the other hand, the corpus-driven approach generates sets of lexical phrases from the corpus and analyses phrasal expressions' use in the corpus (cf. Gray & Biber, 2015). These two approaches are typically used either individually, i.e., a corpus-based method (cf. Moon 1998) or a corpus-driven approach (cf. Altenberg, 1998), or both together (cf. Renouf & Sinclair, 1991).

Several definitions and concepts of phraseology have been introduced above; nevertheless, researchers generally agree that phraseology remains a fuzzy and complex area of language (cf. Altenberg, 1998; Wray, 2002; Granger & Paquot, 2008). Moreover, Stubbs (2007: 181) contends that there is no purely automatic way of identifying phrasal units of meanings. The term *phrase* is used in syntax and phraseology in different ways. The term *phrase* in syntax denotes a grammatical structure that comprises a head and its modifiers, such as noun phrases or verb phrases, that are used within a sentence. However, in phraseology, the term *phrase* frequently refers to expressions that are either fixed or semi-fixed, such as idioms or collocations, and that are influenced by cultural and historical contexts. These expressions express specific meanings. Reflecting the dynamic character of language, this distinction emphasises the interaction between syntax and lexicon.

In this regard, there are language units that are included in formulaic sequences such as internal structures, features of form, irregularities, variability, collocations, functions, meanings, idioms and metaphors, pragmatic meaning, a continuum model (cf. Becker, 1975; Howarth, 1996; Wray, 2002).

These taxonomies are used to identify a broad scope of phraseological perspectives since, as previously mentioned, language is combined into larger units. Table 1 summarises the approaches and classifications of phraseological studies, particularly in relation to structure and form. As can be seen in Table 1, below, two main traditional approaches that have influenced the study of phraseology are represented: the Russian tradition and the Firthian tradition; and it shows the criteria used to identify phrasal expressions (§ 2.3.1).

1. Phraseology	2. Criteria used/Focus	3. Classification of phraseology			
		3.1. Forms			3.2. Discourse functions
		3.1.1. Idiomatic status	3.1.2. Length	3.1.3. Continuous/discontinuous	
1.1. Russian phraseology (cf. Howarth, 1996; Cowie, 1998; Mel'čuk, 1998; Nesselhauf, 2003)	**2.1. Linguistic criteria (§ 2.3.1.1)**	3.1.1.1. Frequent fixed idiomatic expressions	3.1.2.1. Relatively short combinations e.g., 2-3 words	3.1.3.1.Continuous (uninterrupted) sequences	3.2.1. Consideration of discourse functions
	2.1.1. Syntactic: word-like unit/composites				
	2.1.2. Pragmatic: sentence-like unit/formulae				
	2.2. Degree of fixedness	3.1.1.2. Frequent non-idiomatic expressions	3.1.2.2. Extended multi-word squences: 3+ words	3.1.3.2. Discontinuous sequences with variable slots	3.2.2. No consideration of discourse functions
	2.2.1. Free collocation, for example, blow a trumpet				
	2.2.2. Restricted collocation e.g., blow a frame				
	2.2.3. Figurative idiom e.g., blow your own trumpet				
	2.2.4. Pure idiom e.g., blow the gaff				
1.2. Firthian tradition (§ 2.3.1.2)	**2.3. Frequency-based approach (§ 2.3.1.2)**				
1.2.1. Open choice	2.3.1. Collocation/Colligation				
1.2.2. Idiom principle (cf. Firth, 1957; Sinclair, 1991; McEnery & Hardie, 2011)					

Table 1 Traditional Approaches in Phraseology: Key concepts and findings

2.3.1 Two traditional approaches in phraseological studies: concepts and relevant research findings

2.3.1.1 The traditional phraseological approach: Russian phraseology

The concept of a phraseological approach was first mooted in the Soviet Union and Eastern Europe through a movement called Russian phraseology (cf. Howarth, 1996; Mel'čuk, 1998; Cowie, 1998; Granger & Paquot, 2008). Gläser (1988) points out that the concept of this approach can be divided into two units: word-like units and sentence-like units (see also Cowie, 1998: 5). While word-like units employ the syntactic level or below the simple sentence as one way of phraseological categorization, sentence-like units use pragmatics to identify functions, e.g., sayings, catchphrases and conversation formulae (Cowie, 1988). Similarly, the terms *word-like units* and *semantic units* are used differently by different scholars, e.g., *composite unit* (Howarth, 1996), *nomination* (Gläser, 1988), *semantic phraseme* (Mel'čuk, 1998) or *composite* (Cowie, 1998).

Also, a sentence-like unit can be called a functional expression composite (Howarth, 1996), proposition (Gläser, 1988), pragmatic phraseme or pragmateme (Mel'čuk, 1998) or functional expression (Cowie, 1998). In this respect, syntax and pragmatics are used as linguistic criteria to classify phrasal expressions based on frequency (see Table 1, above). In fact, Cowie (1998) separates word combinations into two main types: composites and formulae. Formulae include word combinations that have pragmatic functions, such as: *How are you?* or *Good morning*, whereas composites are collocations that have a syntactic function (see also Howarth, 1996; Cowie, 1998; Nesselhauf, 2005).

As can be seen from Table 1 above, four examples of formulaic sequences are sub-divided into their degree of fixedness, i.e., *blow a trumpet*, *blow a fuse*, *blow your own trumpet*, and *blow the gaff*. Superficially, they have the same syntactic form, consisting of a transitive verb plus an object noun phrase. It is easy to add any object noun phrases that can be followed by the verb *blow*. However, it is unusual to assign the surface meaning to the phrasal expressions *blow the gaff* or *blow your own trumpet*. To assign meanings to formulaic sequences and be able to label them on a continuum from transparent to non-transparent, Nesselhauf (2005: 14) states that the degree of transparency hinges on:

> *...whether the elements of the combination and the combination itself have a literal or non-literal meaning, and commutability refers to whether and to what degree the substitution of the elements of the combination is restricted.*

In this regard, Cowie (1998) proposes the phraseological continuum model: free combinations, restricted collocations, figurative idioms and pure idioms (cf. Howarth, 1996; Cowie, 1998; Nesselhauf, 2005). Formulaic sequences can be assigned meanings by the degree of transparency (see also Cowie, 1998; Nesselhauf, 2005; Granger and Paquot, 2008). Cowie (1998) emphasises the occurrence of words or collocations. That is to say, target words can occur in any phrase in a so-called free collocation or literal sense, i.e., "*blow the trumpet*". In this example, the word blow as a verb of action can occur with a musical instrument as a noun phrase (a trumpet). "*Blow a fuse*" is considered to be a restricted collocation in which substitution of the noun phrase (a fuse) is possible, and at least one element has a non-literal meaning and one a literal meaning; their combination should be transparent (Nesselhauf, 2005: 14). In the figurative sense, "*blow your own trumpet*" conveys a figurative idiom, meaning to tell everyone proud of your achievements. In the figurative meaning, 'substitution of the element is seldom possible', while the combination has a figurative meaning (Nesselhauf, 2005: 14). For a pure idiom (i.e.,*blow the gaff*), Nesselhauf (2005) suggests that substitution of the element and meanings is not possible. Formulaic sequences that fall into pure idiom have a figurative meaning and do not carry a literal sense (cf. Howarth, 1996; Cowie, 1998; Nesselhauf, 2005).

2.3.1.2 The Firthian tradition

While the Russian tradition uses linguistic criteria for the identification of multi-word sequences, i.e., syntactic and semantic restrictions, the Firthian tradition is guided by a distributional or frequency-based approach (cf. Cowie, 1998; Nesselhauf, 2005; Granger & Paquot, 2008), see Table 1 above. The criteria for identifying word combinations involve looking at linguistic meanings: collocations and colligations (cf. Howarth, 1996; Cowie,

1998; Granger & Paquot, 2008). Similarly, the Firthian tradition focuses on the tendencies of words that co-occur, using frequency or distribution as the main criteria.

The development of the phraseology tradition in the Soviet Union and Eastern Europe has been extended to Neo-Firthians,[8] a group which introduced the concept of collocations. Firth (1957: 196) views collocations as the 'abstract tendency of words to co-occur'. He attests to instances of the phenomenon being observable in texts. According to Xiao (2015), one example of a collocation is a word group, such as letter, stamp and post office, or hair, comb and wave. In other words, lexical items are often associated or co-occur with the concepts they refer to in context, and this has been called *coherence collocation*. According to Firth's (1957: 196) notion of collocation, in relation to the term *coherence collocation*, *meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with a conceptual or idea approach to the meaning of words*. Apart from pairs of words, Firth tries to investigate the relationship between words which are separated by longer stretches of text (Xiao, 2015). However, collocational relationships in the textual environment remain unclear. Again, Firth is uncertain about the routine occurrence of words or *habitual collocations.* He questions the concept of frequency of occurrence of collocation, based on his famous examples: *You silly ass!, Don't be such an ass!, What an ass he is!*. It appears that the term *ass* is frequently used together with the other words in the quotes, as evidenced by this example. This has been defined as habitual collocation (Firth, 1957: 11). Nevertheless, Firth does not define the term *frequency*.

Drawing on from Firth's (1951) collocational concept, Sinclair (1991) studies collocations using frequency and statistical methods, such as log-likelihood, t-test and mutual information. Sinclair (1991) uses the term 'neighbourhood collocation' to refer to 'words that co-occur with a word in context' (Scott & Tribble, 2006). Sinclair (1966) notes that Firth's 'coherence collocation' is difficult to measure. Instead, he uses the term 'neighbourhood collocations', which can be retrieved using computational methods and

[8] A traditional British school in corpus linguistics involving a group of scholars, i.e., Halliday, Sinclair, Stubbs, Hoey and Hunston. Regarding corpus data, the corpus-based approach includes the study of lexis, grammar and collocations, which are inseparable (cf. Sinclair, 1987; Gledhill, 2011; McEnery, Xiao, & Tono, 2006; Flowerdew, 2012).

can be measured by statistics. Based on Sinclair's neighbourhood collocation, the frequency-based approach was introduced (cf. Howarth, 1996; Xiao, 2015). Based on frequency, Sinclair (1991) created the model of extended lexical units and their structures: collocation, colligation, semantic preferences and semantic prosody. According to Sinclair (1991), a collocation is a 'relation between the node word and individual word-forms which co-occur frequently with it' (Stubbs, 2007). The relationship between a node word and grammatical categories that frequently co-occur with it is called 'colligation'. The third structure is semantic preference. *Semantic preference* refers to the tendency of certain words (*nodes*) to co-occur with other words that belong to a particular semantic field or category. This relationship highlights how specific words are associated with particular meanings based on their typical collocates. For example, the verb *commit* is frequently found with negative nouns like *crime* or *offence*, demonstrating a semantic preference for negative contexts (McEnery & Hardie, 2011). Similarly, the lexical set of *quantities and sizes*, which includes terms like *number*, *scale*, *amounts*, and *quantities*, reflects the semantic preference of associated words for expressing magnitude. This understanding helps contextualise the usage of words within discourse and their semantic grouping.

Semantic prosody describes the attitudinal or evaluative meaning that a word acquires through its frequent co-occurrence with particular collocates. *Semantic prosody*, first introduced by Sinclair (1991), reflects the positive or negative connotations that arise from habitual word associations. For instance, the verb *cause* is inherently neutral in isolation but frequently appears in negative contexts such as *cause damage*, *cause problems*, or *cause pain*, resulting in a negative semantic prosody. In contrast, *produce* often carries neutral or positive associations, as seen in phrases like *produce results* or *produce food* (Stubbs, 2007; Partington, 2004). These distinctions highlight the impact of semantic preference and prosody on the perception and understanding of words in speech. *Semantic preference* relates to the collocational associations between a word and its semantically related counterparts, but *semantic prosody* emphasises the emotional or evaluative aspects that emerge from their frequent co-occurrence.

Sinclair (1991), working with corpora, described two systems for identifying text behaviour: the open-choice principle and the idiom principle. The open-choice principle, also referred to as the slot-and-filler model, emphasises syntactic relations where language is constructed piece-by-piece, with words freely selected as long as they follow grammatical rules (Sinclair, 1991; Wray, 2002). This model aligns with traditional views of syntax, which focus on the structured arrangement of words and have been studied for millennia. It also relates to Chomsky's (1965) notion of linguistic competence, which views grammatical knowledge as a key trait of proficient language users. While inserting words into grammatical slots may produce structurally correct sentences, this does not always ensure naturalness or idiomatic meaning. This limitation is what the idiom principle addresses, highlighting how fluent speakers often rely on pre-constructed phrases to achieve natural communication. Nevertheless, this method is in contrast to the idiom principle, which emphasises prefabricated language units. The open-choice principle is the basis for non-formulaic elements of language, in which flexibility in word choice is prioritised, as stated by Altenberg (1998). The idiom principle highlights how speakers often rely on pre-constructed expressions stored as units. These expressions range from fully fixed to semi-fixed, showing varying degrees of fixedness depending on context (Brett, 2010; Erman & Warren, 2000). Importantly, Sinclair did not invent the open-choice principle, but described it within corpus linguistics, emphasising its interaction with the idiom principle. He argued that neither principle alone can fully explain language use, as speakers often alternate between constructing language word-by-word and using pre-fabricated phrases, especially in spontaneous speech.

Pawley and Syder (1983) further argue that sentences constructed using the open-choice principle may sound grammatical but might not always feel natural in actual use. Consider the two sentences: *The captain has illuminated the seatbelt sign as an indication that landing is imminent* and *The captain has put the seatbelt sign on, which means we're about to land* (cf. Wray, 2002:13; Wang, 2016). The second sentence demonstrates a more natural-sounding phrasing that aligns with idiomatic usage, illustrating the limitations of the open-choice principle in creating language that feels authentic in practical contexts.

Wray (2002) critically argues that, while the grammar of the first sentence is acceptable, it lacks the naturalness of the second sentence. A native speaker would likely prefer the second sentence because it aligns with native-like selection and conventional usage, which go beyond grammatical correctness to incorporate idiomatic and prefabricated units of language (cf. Wray, 2002). In the same vein, language cannot be fully understood by grammar rules alone because much of it is made up of larger units called prefabricated units, which follow the established rules of syntax and contribute to the structure of sentences in systematic ways. Prefabricated units, including fixed expressions and idioms, are not always explained through the application of grammar rules alone. For instance, the importance of idioms such as *kick the bucket* does not come from the meaning of each individual word, but rather from the whole phrase. The open-choice principle is used to explain the process by which words can be freely selected to form sentences, provided that they follow to the rules of grammar. Nevertheless, this principle is unable to consistently clarify formulaic language, as it is relied on preconstructed phrases, which are explained by the idiom principle (Altenberg, 1998; Wray, 2002). Sinclair (1991:114) explains that the idiom principle is the main way for understanding a large number of texts, as they are full of formulaic and prefabricated expressions. He states that:

> ...the first mode to be applied is the idiom principle, since most of the text will be interpretable by this principle. Whenever there is good reason, the interpretive process switches to the open-choice principle, and quickly back again. Lexical choices which are unexpected in their environment will presumably occasion a switch... .

Thus, the idiom principle refers to formulaic elements of language, like idioms, collocations, and fixed phrases (cf. Altenberg, 1998: 101; Wray, 2002; Granger & Paquot, 2008). These formulaic units help language users communicate more easily by relying on familiar patterns. In contrast, the open-choice principle allows flexibility and creativity when language users create sentences that do not rely on prefabricated phrases. Similarly, Pawley and Syder's (1983) view of the language of native speakers is that there are

substantial numbers of formulaic sequences and chunks that are used by them, such as *How are you?* (see also Prodromou, 2008; Paquot & Granger, 2012). Thus, formulaic sequences play an important role in language learning. In the same way, formulaic sequence is one of the units in the language system. So, the best way to solve this issue of multi-word units is to define and identify fixed multi-word units, and analyse the discourse functions that these multi-word units perform (cf. Weinert, 1995; Gray & Biber, 2015).

As shown in Table 1, formulaic sequences can be classified according to their form and function. Meanings can be categorised in a continuum model, from the most transparent or free collocations to pure idioms (§ 2.3.1.1 for a discussion of the continuum model). To be able to identify communicative purposes in word combinations, discourse functions can be defined, depending mostly on the nature of the phraseology (cf. Nesselhauf, 2004; Gray & Biber, 2015). Some of the phrasal expressions labelled are focused on their forms (3.1), excluding their functions (3.2), as shown in Table 1, above. Lexical sequences can be grouped into fixed idiomatic expressions or non-idiomatic sequences of frequent words (cf. Howarth, 1996; Wray, 2002; Gray & Biber, 2015). In terms of the length of lexical sequences, there are short combinations of two or three words, or extended multi-word sequences comprising more than three words. Regarding form, lexical sequences can be divided into continuous or uninterrupted (e.g., the number of) or discontinuous with variable slots (e.g.,*in the * of*). Lexical sequences can be categorised into functions, such as pragmatic functions (cf. Nattinger & Decarrico, 1992). Research findings that use phraseological methods for labelling, identifying and classifying formulaic sequences are represented in Section 2.3.1.3, as shown in Table 2, below.

Study, -ies	1. Phraseological label(s)	2. Classification of phraseology			
		2.1. Forms			2.2. Function
		2.1.1. Idiomatic status	2.1.2. Length	2.1.3. Continuous/ discontinuous	
Moon, 1998	Phrasal lexemes	idiomatic/frequent non-idiomatic	short: 2–3 words/extended: 3+ words	fixed/semi-fixed	consideration of discourse function
Altenberg, 1998	recurrent word combinations	frequent non-idiomatic	short: 2–3 words	continuous	consideration of discourse function
Nattinger & Decarrico, 1992; Granger, 1998	lexical phrases, prefabricated patterns	idiomatic	extended: 3+ words	discontinuous	consideration of discourse function
Biber et al., 1999; Biber & Barbieri, 2007; Hyland, 2008; Cortes, 2013	lexical bundles, words cluster, formulas	frequent non-idiomatic	extended: 3+ words	continuous	consideration of discourse function
Nesi & Basturkmen, 2006; Csomay, 2013	lexical bundles	frequent non-idiomatic	extended: 3+ words	continuous	consideration of discourse function
Liu, 2012	multi-word constructions	non-idiomatic but frequent	short: 2–3 words/extended: 3+ words	continuous	consideration of discourse function
Eeg-Olofsson & Altenberg, 1994	discontinuous frames	frequent non-idiomatic	short: 2–3 words	discontinuous	no consideration of discourse function
Renouf & Sinclair, 1991	collocational frameworks	frequent non-idiomatic	short: 2–3 words/extended: 3+ words	discontinuous	no consideration of discourse function
Biber, 2009	lexical bundles and lexical frames	frequent non-idiomatic	extended: 3+ words	continuous/ discontinuous	no consideration of discourse function
Römer, 2010; Garner, 2016; Cunningham, 2017	phrasal frames (p-frames)	frequent non-idiomatic	extended: 3+ words	continuous/ discontinuous	no consideration of discourse functions
Gray & Biber, 2013	lexical frames	frequent non-idiomatic	extended: 3+ words	continuous/ discontinuous	no consideration of discourse function
Simpson & Mendis, 2003	idioms	fixed idiomatic	short: 2–3 words/extended: 3+ words	continuous	consideration of discourse function
Simpson-Vlach & Ellis, 2010	formulas	non-idiomatic	extended: 3+ words	continuous	consideration of discourse function
Aijmer, 2009; Grigaliūnienė & Juknevičienė, 2013	formulaic sequences	frequent non-idiomatic	extended: 3+ words	continuous	consideration of discourse function
Aston, 2015	phraseology	frequent non-idiomatic	extended: 3+ words	continuous	no consideration of discourse function

Table 2 Research findings for extended lexical sequences (adapted from Gray & Biber, 2015: 134)

2.3.1.3 Previous research on phraseological perspectives

Since the present study adopts one of the units of phraseological identification, lexical bundles, in an academic register, i.e., engineering research articles, previous research on operationalising lexical sequences and their measurement with help from corpus analysis will now be outlined. As previously mentioned, studies of lexical phrases have been used in both corpus-based studies and corpus-driven studies (§ 2.2 for a discussion of the two approaches). There are many studies that generate a full set of lexical phrases (i.e., corpus-driven studies) using a computer program (e.g., AntConc, ConcGram, WordSmith Tools), and then their frequency of appearance in a corpus is identified. Frequency criteria are used to determine the frequency with which lexical sequences occur in the corpus and to quantify the significance of those sequences, thereby identifying which sequences require further investigation (cf. Biber et al., 1999; Conrad & Biber, 2005; Flowerdew, 2012). The relationship between frequency and noteworthiness lies in the use of frequency as a marker of significance in language, consistent with Biber et al. (1999) and Conrad & Biber (2005). Frequent lexical sequences are more likely to be recognised and processed efficiently, making them important for both pedagogical and linguistic analysis. Noteworthiness is not solely determined by frequency; contextual relevance, discourse functions, and cognitive processing also play a substantial role in the selection of lexical sequences. For example, a sequence that is highly frequent may not be considered noteworthy if it is irrelevant in a particular context or does not fulfil a clear communicative purpose.

Regarding the length of lexical sequences, 2-word sequences (collocations), and sequences of three words or more (extended collocations) in length are normally investigated (cf. Moon, 1998; Altenberg, 1998; Xiao, 2015). To be able to manage the length of lexical sequences, 3-word and 4-word sequences are commonly studied because at this length it is easier to get rid of fragmentary sequences (cf. Biber et al., 2004; Hyland, 2008; Chen & Baker, 2010). Fragmentary sequences refer to incomplete or partial lexical sequences that do not form meaningful or coherent expressions. These sequences often arise from errors, incomplete phrases, or arbitrary combinations of words that do not reflect standard usage. For example, in a corpus analysis, sequences such as *in the* or *of this* may

be identified as fragmentary because they lack the context or structure needed to convey a complete idea. By focusing on 3-word and 4-word sequences, researchers can filter out these incomplete units and focus on meaningful patterns, such as *in the context of* or *at the end of*. For practical reasons, some specific lexical sequences that belong to particular types of syntactic patterns may be excluded from the selection, such as phrasal verbs and compound nouns (cf. Moon, 1998). In this regard, operationalising phraseological units, their measurement (e.g., frequency, an association of lexical sequences) and which corpus methods are used depends on the purpose of individual research (Howarth, 1996; Wray, 2002). Based on frequency criteria, one of the pioneer studies of lexical phrases can be seen in Salem (1987). He identified lexical phrases in a corpus of French government texts based on frequency criteria (cf. Biber et al., 2004; Gray & Biber, 2015).

Based on the Russian tradition, Moon (1998) employed an 18-million word corpus of British English in her analysis of fixed and semi-fixed phrases[9] called phrasal lexemes (see Table 2 above). Concerning statistical calculations, frequency range was also used to identify the number of phrasal lexemes which occurred in each range (see Moon, 1998: 83). In terms of distribution, she tests hypotheses about correlations between frequency, form, type and function. In the same vein, linguistic sequences are analysed using linguistic criteria: lexico-grammar, syntax and pragmatics (cf. Moon, 1998b: 84). However, she excludes phrasal verbs, compound nouns, adjectives and verbs and incorporates fewer frequent lexical sequences to gather idioms in her study. Another study based on the investigation of continuous lexical sequences of spoken English is the London-Lund Corpus (see Svartvik, 1990). Observations on pre-patterned expressions, including idioms, fixed phrases and collocations, can be found in Altenberg (1998). Regarding the most common expressions, the study reveals that a wide range of linguistic elements can be found, i.e., grammatical structures, complete phrases, incomplete phrases, core sentences, clause elements (see Altenberg, 1998). Based on frequency and distribution criteria, Moon

[9] The terms fixed and semi-fixed phrases in Moon's (1998) study include phraseological chunks, such as idioms, metaphors, proverbs, sayings (see Moon, 1998a:19–25 for detailed definitions of these types).

(1998) introduces the notion of lexicogrammatical frames, which is similar to the study of collocational frameworks in Altenberg (1998) and Renouf & Sinclair (1991).

Since formulaic sequences have form and function, Nattinger & Decarrico (1992) classified lexical phrases of three words or more according to the internal structure of phraseological units: polywords, phrasal constraints and sentence builders. Polywords are multiword expressions that function as single units in communication. Examples include idiomatic expressions such as *pull someone's leg*, common phrases like *see you later*, and fillers such as *you know*. Phrasal constraints refer to fixed structures within multiword expressions, such as idioms or collocations, that maintain their integrity and meaning. For instance, the idiom *kick the bucket* cannot be altered without losing its meaning. Sentence builders, on the other hand, provide learners with pre-formed language chunks that aid in constructing sentences. Examples include conversational formulae such as *thank you very much* or *how are you?* and collocations like *face a challenge* or *make a decision*. These classifications highlight the diverse roles formulaic sequences play in communication and language learning.

By comparing the functions of prefabricated phrases in three different languages, they showed that these internal structures have three main functions: social interaction, necessary topics, and discourse devices. In social interactions, these phrases provide a shared linguistic framework that enhances mutual understanding (Gülich, 2008). For example, expressions like *how are you?* or *nice to meet you* help establish rapport and navigate social norms. Necessary topics are addressed using ready-made expressions that streamline communication, such as *can you help me with this?* or *I would like to order a coffee*. These phrases are particularly useful in routine or predictable contexts, supporting effective communication. As discourse devices, prefabricated phrases help structure conversations and ensure coherence (Gülich, 2008). Phrases like *on the other hand*, *in conclusion*, or *for example* signal transitions, emphasise points, or introduce new topics, thus maintaining interactional flow. These examples demonstrate how prefabricated phrases serve as practical tools for facilitating the social, functional, and structural aspects

of communication across languages. The findings show the same patterns across three languages.

Phraseological methods of analysis are used to identify learners' lexical sequences as used in written and spoken language. Granger (1998) compares the use of prefabricated patterns by native speakers and learners in two corpora: the International Corpus of Learner English (ICLE) and the Louvain essays corpus. She identifies collocational patterns and formulae, along with their functions. The study suggests that learners underuse native-like collocations and typical word combinations (Granger, 1998: 152). While learners share similar numbers of passive frames with native speakers, they overuse active structures. These two studies imply the need for an emphasis on teaching the principles of lexical sequences, i.e.,form and function (Wray, 2002). Similarly, Grigaliūnienė & Juknevičienė (2013) identify formulaic sequences from continuous 2–5 word sequences in learners' written and spoken language. They find that formulaic sequences are important in producing spontaneous speech and learners rely heavily on them. Their study also suggests that there are overlapping core sets of formulaic sequences (e.g.,*at the same time, as well as the*), and that most of them are discourse-organising expressions. The formulaic sequences that are used by students are semantically transparent and their formulaic nature depends on external or contextual factors. Based on the LINDSEI corpus, Aijmer (2009) studies the functions of formulaic sequences (i.e.,*if I don't know* and *(I) dunno*) to see if selected formulaic sequences are multi-functional. The findings show that *I don't know* is usually used to mark the end of a topic or to express uncertainty. However, learners use *I don't know* as a speech management signal, rather than for other functions. Maybe it is because learners do not have time to plan their sentences as well as native speakers do (cf. Aijmer, 2009).

Drawing on a frequency-based approach, lexical bundles are traditionally considered only to be continuous sequences of three words. Biber et al. (1999) analyse continuous lexical bundles that appear frequently in written and spoken language across various genres. In their analysis, they exclude idioms because lexical bundles are typically identified based on frequency and function, not fixed idiomatic meaning. Unlike idioms,

lexical bundles often serve structural or discourse-organising functions and are not necessarily semantically opaque. Therefore, idioms are treated as a separate category in their study. Most work uses frequency as the criterion to identify lexical sequences, as well as in analysing their internal structure and functions. Based on lexical structure and function, Hyland (2008) and Cortes (2013) determine lexical bundles' functions in academic work. Hyland (2008) uses a corpus made up of his students' academic writing in the fields of engineering, biology, business and applied linguistics. Hyland (2008) separates functions into three categories: participant-oriented, text-oriented and research-oriented. Cortes (2013) generates the lexical bundles that occur in the introduction part of research articles. In particular, she found there are functions tied to specific moves or steps. Drawing on lexical bundles' forms and functional taxonomies from Biber et al. (1999), Nesi & Basturkmen (2006) identify the structures and functions of lexical bundles that are used in classroom lectures (spoken language). Similarly, Csomay (2013) investigates lexical bundles' forms and functions used in the classroom. The results of his study suggest that there is a strong relationship between grammar and lexis on the discourse level. Drawing on Biber et al.'s (2004) and Conrad & Biber's (2005) criteria for lexical bundles' measurement, Liu (2012) develops the 228 most common multi-word constructions (MWCs) in both COCA and the BNC. He groups MWCs into three bands, and each band is allocated to a specified range. Then, he identifies the structures and functions of MWCs. Based on their frequency and semantic functions, these generated MWCs prove useful in the study of general academic writing. Also, it is suggested that filled constructions (e.g.,idioms) are rarely used in such writing.

Research on discontinuous sequences has influenced the study of phraseology, such as Eeg-Olofsson & Altenberg, (1994), Renouf and Sinclair (1991) and Römer (2010). A collocational framework is considered to be a 'linguistic skeleton of full bodied phrases' (cf. Sinclair, 2008: 409). The methodology of these studies involves investigating the number of distinct fillers that occur in variable slots, using statistical measures such as the type/token ratio to assess the internal fixedness of each frame. Renouf and Sinclair (1991) are interested in identifying a collocational framework for non-continuous sequences and

fixed lexical frames combined with open slots (see the same methods in Moon, 1998). They use the term 'collocational frameworks' or pairs of words separated by one intervening word (e.g., *a*+_____+ *of*). The study reveals that common fillers occur in each frame (cf. Conrad & Biber, 2005). However, the sequences in this study are not syntactically complete (Biber & Conrad, 1999).

Drawing on the work of Renouf and Sinclair (1991), Eeg-Olofsson & Altenberg (1994) investigate discontinuous frames (e.g., *the_of, I_know*) in the London-Lund Corpus (Svartvik, 1990). The 1,000 most common frames (e.g., *the_of*) are analysed for type-token ratios of filler words. Via computational and statistical analysis techniques, they use MI scores[10] to identify how predictable frame fillers are. The highest-frequency frames are combined with function words. There is no functional analysis in this study. While medium-frequency frames tend to refer to abstract or general lexical words, low-frequency frames have more lexical words with fillers that are less variable (cf. Eeg-Olofsson & Altenberg, 1994; Altenberg, 1998; Gray & Biber, 2015).

Biber (2009) combines corpus-driven studies and corpus-based studies in identifying lexical bundles in conversation and academic prose. He first focuses on lexical sequences from a corpus and then identifies the degree of fixedness of lexical sequences. He further investigates the nature of extended collocations in conversation and academic prose. The results reveal that academic discourse mainly relies on frames with intervening variable slots. Fixed sequences are filled by function words. Building on Biber (2009), the identification of a full set of discontinuous lexical sequences or p-frames can be found in Römer (2010) and Gray and Biber (2013). In particular, Gray and Biber (2013) classify three structures of phrasal frames. The first one is a *verb-based frame,* consisting of a frame which contains one or more modal, auxiliary or main verbs (e.g., *can speak * languages*, *is * my house, I * like to*). Second, *frames with other content words* include frames that contain one or more nouns, adjectives or adverbs, but no verbs (e.g., *I * from Germany, an absolutely* job, the * stage of*). The last frame is a *function word frame* that consists of

[10] Mutual Information (MI) score is a statistical measure used in corpus linguistics to assess the strength of association between two co-occurring words or phrases. Higher MI scores indicate stronger associations (see McEnery & Wilson, 2001; Biber, 2009; McEnery & Hardie, 2011; Cortes, 2015).

only function words, such as prepositions, determiners, pronouns etc. (*my * and I, the * in the, a * for a*).

Drawing on the phrasal frames in Römer's (2010) and Gray and Biber's (2013) studies, the notion of phrasal frames (p-frames) has been used to identify lexical sequences and their function in learner writing. Based on a corpus-driven approach, Garner (2016) examines p-frames' structure and their functions as used by L1 German learners of English as a Foreign Language at five different proficiency levels represented in the EF-Cambridge Open Language Database (EFCAMDAT). Garner (2016) found that the use of p-frames by higher proficiency learners is more variable, less predictable and more functionally complex. A comparative study of p-frames between the two corpora is also included. In the same vein, Cunningham (2017) compares p-frames and their function in a mathematical research articles corpus with the academic section of the Corpus of Contemporary American English (COCA). Since mathematical texts include multiple proofs, where prior knowledge is expected, p-frames can be filled out using a restricted set of lexical words. For example, the p-frame i.e., *this * the proof of* can be followed by one of a restricted set of synonymous fillers, such as *completes, concludes, finishes,* in order to indicate a proof's completion in mathematics (Cunningham, 2017: 77).

Simpson & Mendis (2003) focus on idioms to help students learn unfamiliar idioms, based on the Michigan Corpus of Academic Spoken English (MICASE). An idiom consists of a group of words that occur in a more or less fixed sequence. The meanings of idioms are not transparent (§ 2.3.1.1 for a discussion of the Russian approach). In the context of linguistic transparency, the Russian approach to idioms includes a study of the compositional nature and regularities of idioms within the context of syntactic theory. The potential for idioms to transparently reflect the morphosyntactic features of their components can be seen by this approach, which contrasts with the traditional focus on idiomatic irregularities. For instance, the compositional properties of Russian idioms are examined, indicating that the morphosyntactic characteristics of their elements can be used to understand the idioms (Tronenko, 2003). This viewpoint enhances comprehension of the syntactic flexibility and semantic complexity of Russian phrasal idioms (Tronenko, 2001).

The use of a triangulating corpus and psycholinguistic research can be seen in Simpson-Vlach & Ellis (2010). They identify not only the most common formulaic expressions but also include low frequency lexical sequences in their analysis. The formulaic expressions found are in both core and overlapping patterns. In terms of validity, they employ Mutual Information (MI) to judge the salience of lexical sequences. The major contribution of this study is its enhancement of a method for identifying LBs, since lexical sequences are vital in studying language, in particular English (O'Donnell et. al., 2015). To increase learners' awareness and repertoires of phraseological sequences in spoken language, Aston (2015) identifies continuous lexical sequences from a Technology, Entertainment, Design (TED) talks corpus and creates listening activities to teach from a lexical sequences list. This research developed listening activities based on continuous lexical sequences from TED speeches, drawing influence from Aston (2015), who examines the use of these sequences for developing listening tasks for teaching vocabulary (see Aston, 2015). By focusing on continuous lexical sequences within the TED talk corpus, participants noticed the meanings of lexical bundles using their stress patterns and prosodic structures. Listening to these sequences in context enabled learners to understand the role of these lexical bundles in natural speech, highlighting the significance of emphasis, rhythm, and intonation in message conveying. This method helps learners in identifying the auditory features of the language with its contextual use, hence improving vocabulary acquisition and retention.

Apart from the benefits of phraseology in learning English or other languages, there are studies of the effects of phraseology in L2 English learning. An investigation of the phraseological approach in students' writing can be seen in Osborne (2008). As previously presented in Section 2.3.1, phraseological units include syntactic and lexical elements (cf. Howarth, 1996; Wray, 2002). The meanings of lexical sequences are associated with linguistic units, which help maintain linguistic boundaries and influence written production (cf. Howarth, 1996; Wray, 2002; Osborne, 2015). Some errors may occur and persist while learners produce pieces of writing, such as pluralised adjectives, pluralised mass nouns, third person -s. In Osborne (2008), the three kinds of phraseological effects in written productions are categorised as blending, bonding and burying. Blending occurs when

learners deal with elements that are combined into larger units of language. Osborne (2015: 81) investigates the term bonding in relation to the 'the learners' lexicon and grammar, and lexical or grammatical elements, e.g., with adverb + verb collocations such as 'follow everything blindly'. In this connection, Osborne (2015: 81) describes burying, which occurs when an 'element which is embedded inside larger units may become less salient, and so lose grammatical features that they would normally be expected to carry'.

As discussed above, phraseology can have many forms, and they are defined differently. Since it is difficult to identify phraseological units, frequency is used as the criterion for the identification of formulaic sequences (Granger & Paquot, 2008; Flowerdew, 2012). Using a frequency-based approach, the Firthian tradition focuses on collocations and investigates the tendency of words to occur with node words (target words) in a corpus within a certain span, i.e.,1–3 words from the left and the right (cf. McEnery & Hardie, 2011; Xiao, 2015). Also, by looking at colligation, grammatical patterns can be revealed. Based on frequency, the Russian tradition uses syntactic (word-like units) and pragmatic (sentence-like units/ formulae) units as the main criteria to identify formulaic sequences (cf. Cowie, 1998, Mel'čuk, 1998). Regarding the classification of lexical sequences, the Russian tradition ranges them on a continuum which goes from free combinations to pure idioms (cf. Howarth, 1996; Cowie, 1998). This section also reviews research related to phraseological methods: frequency and threshold, nature of phraseological labels and classification of phraseological labels. The next section is a review of phraseological units called lexical bundles, including concepts, structures and discourse functions.

2.4 Lexical bundles: concepts, previous studies, structure and discourse functions

Before analysing lexical bundles in engineering research articles (see Chapter 3), it is crucial to understand related concepts and methods of identification. In sub-section 2.4.1, key concepts around lexical bundles, including definitions and characteristics, are defined. Their structures and functions are explained in sub-sections 2.4.2 and 2.4.3. Then, previous studies of lexical bundles are introduced in sub-section 2.4.4. All the above elements are essential components of lexical bundle identification. In particular, these components of

lexical bundles contribute to the underlying interpretation of the structure and functions of lexical bundles used in engineering research articles (see Chapter 3).

2.4.1 Concepts of lexical bundles: definitions and characteristics of lexical bundles

As previously noted in section 2.3, phraseology recognises different types of units including collocations, extended collocations, idioms and fixed phrases (cf. Wray, 2002, 2009; Meunier & Granger, 2007; Granger & Paquot, 2008). Lexical bundles are considered to be 'another type of phraseological unit which is identified by purely 'frequency-driven' means' (Flowerdew, 2012: 71). In the same vein, to be able to recognise or identify lexical bundles, the statistical analysis of a language corpus is involved (cf. Salem, 1987; Cortes, 2015). Drawing on the traditional Firthian perspective, the study of lexical bundles is the study of extended collocations (cf. Biber et al., 1999; Cortes, 2015). Different scholars have used and defined various terms for multi-word sequences, including collocations, lexical phrases, formulas, prefabricated expressions, idioms, n-grams, lexical bundles, frames, and collocational frameworks. These terms reflect the diverse ways researchers approach and categorise recurring word patterns in language. Lexical bundles are identified using a fully corpus-driven methodology which gives priority to frequency in identifying lexical sequences (Altenberg, 1998; Biber et al., 2004; Conrad & Biber, 2005). Since lexical bundles are seen as units of language components (cf. the idiom principle, Sinclair, 1991), each unit has a function. In this sense, bundles are seen as 'building blocks of discourse, associated with basic communicative functions' (Biber, Conrad, & Cortes, 2004: 400). Examples of lexical bundles found in conversation are expressions such as, *I don't know what to do, you won't be able to, do you want to go*. In academic prose, lexical bundles are word sequences like *as shown in figure, on the other hand*.

To qualify as a lexical bundle, it is essential to identify both its defining characteristics and its discourse function (cf. Biber et al., 1999; Conrad, & Cortes, 2004; Gray, 2016). According to Conrad & Cortes (2004) & Gray (2016), lexical bundles can be distinguished from other phraseological units. As such, the identification of lexical bundles is strictly based on the following criteria in analyses: frequency or cut-off point, dispersion

range, length of lexical bundles, and transparency of lexical bundles, as detailed below (cf. Conrad & Biber, 2005; Biber & Barbieri, 2007):

- Commonality is the key concept of the lexical bundles approach. To be able to consider a lexical bundle common, a criterion for the frequency of occurrence is justified. Based on Biber and Conrad (1999), the frequency cut-off set should be at least 10 per million words in the register. This threshold ensures that lexical bundles are statistically significant and frequent enough to warrant analysis. The cut-off point varies depending on the length of the lexical bundles. For instance, shorter bundles, such as four-word bundles, are set to have a higher cut-off point compared to longer bundles, such as five- and six-word bundles. This difference can be attributed to the unique structural and functional characteristics of four-word bundles, as supported by various studies on lexical bundles (cf. Altenberg, 1993).
- The second priority is the dispersion range, which refers to occurrences of lexical bundles in different elements of a corpus to guard against idiosyncrasies. For example, lexical bundles must be used by multiple speakers or occur in at least five different texts in a target corpus. In this study, lexical bundles are identified based on their frequent occurrence across texts and their presence in at least five different texts in the target corpus. This decision ensures that the identified bundles represent widely-used patterns rather than idiosyncratic expressions. Various studies emphasise the importance of frequency and distribution in identifying lexical bundles, highlighting their role in different contexts and among diverse groups of speakers (cf. Jalilifar, Ghoreishi & Emam Roodband, 2016).)
- Regarding the length of lexical bundles, there are 3-, 4-, 5- and 6+-word bundles, and they must be continuous. Shorter bundles are often incorporated into more than one longer lexical bundle, e.g., *I don't think* is also part of *well I don't think* or *well I don't think so*. The typical length is 3-word and 4-word bundles, as these are manageable and can get rid of fragmentary words, which limits the scope of the study. Some studies categorise 3-, 4-, 5- or 6+-word bundles as core recurrent sequences with optional

extensions, such as the bundle *I don't think,* which can be extended by an optional extension, e.g., *well, oh* (cf. Altenberg, 1993).

- Lexical bundles are normally incomplete structural units, and they often bridge two units: clausal units and phrasal units.
- Regarding the idiomatic meaning of lexical bundles, they have non-idiomatic meanings and are not particularly perceptually salient.

2.4.2 Structure of lexical bundles

Since lexical bundles are involved in the study of words' fixedness and the grammatical functions they are related to, the internal structure of lexical bundles should be identified (cf. Biber et al., 1999; Cortes, 2015). Biber et al. (2004: 380–381); in particular, note that the internal structure of lexical bundles is usually not complete units, they have strong grammatical correlates. In this study, to be able to identify the structures of lexical bundles in research articles, the grammatical patterns in Biber et al. (1999: 1015–1024) are used.

There are 12 structural categories: noun phrase with *of*-phrase fragment, noun phrase with other post-modifier fragment, prepositional phrase with embedded *of*-phrase fragment, other prepositional phrase fragment, anticipatory *it* plus verb phrase/ adjective phrase, passive verb plus prepositional phrase fragment, copula be plus noun phrase/ adjective phrase, (verb phrase+) plus that-clause fragment, (verb/adjective+) plus to-clause fragment, adverbial clause fragment, pronoun/ noun phrase plus be plus (+...), and other expressions.

The fact is that lexical bundles comprise phrases, clauses and clausal fragments (Cortes, 2008; Gray, 2016). In written academic genres, lexical bundles are often found to be phrasal, i.e.,combined with fragments of nouns or prepositional phrases. In contrast, lexical bundles in everyday conversation and other spoken registers are mainly clausal, i.e.,combinations of subjects and verbs. For manageability, Biber, Conrad, and Cortes (2004) regrouped the 12 traditional structures into Noun-Phrase (NP) based, Preposition Phrase (PP) based, Verb Phrase (VP) based, and clause (C). The 12 structures are identified in three main categories: verb phrase fragments, dependent clause fragments, noun phrase

and prepositional phrase fragments (cf. Biber, Conrad, & Cortes, 2004; Cortes, 2015; Gray, 2016).

Structural types of bundles	Description	Example bundles
1. Verb phrase fragments	Incorporate fragments of verb phrases, including subject pronouns followed by a verb phrase, the beginning of a verb phrase, and question fragments	*is going to be* *can be used* *as shown in figure* *can be used to*
2. Dependent clause fragments	Include both verb phrase fragments and components of dependent clauses (e.g., complement clauses)	*if you don't want* *not be able to* *I think you should* *you need to get*
3. Noun phrase and prepositional phrase fragments	Consist of noun phrases, often with a head noun and the start of a post-modifier (commonly a prepositional phrase, but also relative or complement clauses)	*at the end of the* *in the context of* *the best way to* *the way in which*

Table 3 Structure of lexical bundles (adapted from Biber, Conrad, & Cortes, 2004)

In this connection, Biber et al. (2004: 380–381) identify three major categories of lexical bundles. As can be seen in Table 3, above, the first structure is composed of verb phrase fragments (VPs). These bundles begin with a subject followed by a verb phrase fragment, or begin with a discourse marker followed by a verb phrase fragment, or start directly with a verb phrase. Examples of these types of expressions are *is going to be*, *can be used to*, *as shown in the figure*.

The second type of lexical bundle is dependent clause fragments (DCFs). Dependent clause fragments typically include a subject and a verb but lack a complete thought or independent clause. For example, *if you want to* and *I want you to* are examples of dependent clause fragments. These bundles have an incomplete clause structure and are dependent on additional information to form a complete sentence.

The third type of lexical bundle incorporates noun phrase and prepositional phrase fragments. In this type, the structural correlate of lexical bundles is made up of noun phrases or prepositional phrases that start the bundle, followed by other noun or prepositional phrase fragments, as in, for example, *the end of the*, *in the context of*, *the way in which*.

2.4.3 Functions of lexical bundles

Not only is the internal structure of lexical bundles studied, but also their communicative purpose (cf. Biber, et al., 2004; Biber, 2006). Gray and Biber (2015) suggest that the best way to investigate lexical bundles is to look at the discourse functions that lexical bundles perform (Gray & Biber, 2015b). There are two main functional taxonomies that are relevant to lexical bundles studies in academic registers in spoken and written language. As illustrated in Table 4, below, the first one is from Biber's classification (see also Biber et al., 2004; Biber, 2006). He classifies lexical bundles into three main groups: stance expressions, discourse organisers and referential expressions, based on academic registers in spoken and written language (cf. Biber, et al, 2004; Biber, 2006). Drawing on the functional taxonomy of Biber, Conrad, & Cortes (2004), Hyland (2008) categorises the functional framework into research-oriented, text-oriented and participant-oriented.

The lexical bundles from Biber's classification are derived from a broader corpus of spoken and written registers which included casual conversation, textbooks, course packs, service encounters, institutional texts and so on (cf. Biber, et al., 1999; Biber, et al., 2004). The first function is referred to as stance expressions. The lexical bundles in this group signify the writer's attitude, judgement and perspective in terms of certainty or uncertainty, and proposition or ability (cf. Biber et al., 1999; 2003). As indicated in Table 4, below, lexical bundles are found in stance expressions, such as *there is nothing wrong, it depends on the, you need to get*.

Discourse functions of bundles	Description	Example bundles
1. Stance expressions	Indicate epistemic, attitudinal, modal or evaluative assessments, including assessments of certainty or likelihood, desire, obligation/ directive, intention/ prediction, and ability	*there is nothing wrong* *I think it is* *if you want to* *it depends on the* *you need to get*
2. Discourse organisers	Signal relationships between previous and forthcoming discourse, by introducing topics, stating focus or elaborating/ clarifying a topic	*I would like to* *on the other hand* *nothing to do with*
3. Referential expressions	Reference physical, abstract or textual entities, often to identify/ focus on that entity, indicate imprecision, or detail attributes such as quantity, framing, time or place	*for the sake of* *the two of you* *end of the day* *is one of the* *the number of*

Table 4 Discourse function of lexical bundles (adapted from Biber, Conrad, & Cortes, 2004)

The second function found is discourse organisers. The lexical bundles in this function help to compose the structure of a text. For example, the lexical bundle *I would like to* is typically used in introducing a topic, such as in the sentence *I would like to talk about X.* However, in other contexts, the same bundle can serve a different function. For instance, *I would like to* can also function as a stance bundle when expressing a desire or intention, as in *I would like to go there.* This demonstrates how the same form can have different functions depending on the context in which it is used.

Furthermore, expressions which indicate imprecision are also included in this category. For example, *is one of the* suggests vagueness or non-specificity by not fully defining the subject or providing exact details. These bundles are typically helpful when making arguments, regulating conversation, and expressing implicit meanings. The function of such expressions can vary across disciplines and contexts, reflecting the underlying research paradigms and discourse strategies. Lexical bundles, such as *is one of the*, are used differently across various disciplines. In medicine, text-oriented bundles are more prevalent, indicating a structured discourse approach; however, their usage is less common in psychology (Choi, 2015). These bundles facilitate the management of imprecision and the construction of complex arguments, with their function and frequency influenced by discipline, language, and genre.

Similarly, some lexical bundles can be multi-functional**.** For example, the bundle *at the end of* can refer to both place and time, as in the bundle *at the end of this paper***.** In addition, the lexical bundles in this category are found to be more personal, referential and directive

Drawing on lexical bundles' functions in Biber, Conrad, & Cortes **(**2004**)** and Biber **(**2006**)**, Hyland **(**2008**)** developed a functional taxonomy that emphasises research-focused genres, and these usually refer to writers and readers in texts**.** In his analysis of a large corpus of academic writing of graduate students, i.e., research articles, PhD dissertations and MA/MSc theses, Hyland **(**2008**)** identifies three functions of lexical bundles**:** research-oriented, text-oriented and participant-oriented**.**

As can be seen in Table 5, below, the research-oriented function is sub-categorised into five aspects**:** *location, procedure, quantification, description* and *topic***.** The purpose

of the research-oriented function is to 'help writers to structure their activities and experiences of the real world' (cf. Hyland, 2008: 13). The lexical bundles that help to map out texts' organisation fall into the location aspect, as they help to indicate time and place, such *as at the beginning of, in the present study*. Procedure is used to refer to the process of research activities such as *the use of the, the purpose of the*. The lexical bundles that indicate quantification include examples such as *the magnitude of the, a wide range of*. The description of materials or activities group includes example bundles such as *the size of the, the structure of the.* Lexical bundles that refer to a specific topic in research include examples such as *in the Hong Kong, the currency board system.*

Discourse functions of bundles	Decription	Example bundles
1. Research-oriented	1. Location	*at the beginning of, in the present study*
	2. Procedure	*the use of the, the purpose of the*
	3. Quantification	*the magnitude of the, a wide range of*
	4. Description	*the size of the, the structure of the*
	5. Topic	*in the Hong Kong, the currency board system*
2. Text-oriented	1. Transition signal	*on the other hand, in addition to the*
	2. Resultative signal	*it was found that, these results suggest that*
	3. Structuring signal	*in the present study, in the next section*
	4. Framing signal	*in the case of, with respect to the*
3. Participant-oriented	1. Stance feature	*are likely to be, may be due to*
	2. Engagement feature	*it should be noted that, as can be seen*

Table 5 Discourse functions of Hyland's (2008) lexical bundles (adapted from Hyland, 2008)

For instance, the phrase *the size of the* might be categorised under description since it conveys an aspect or characteristic of something, but it could also be argued to belong under quantification, as it refers to a measurable attribute (size). This highlights the potential ambiguity of certain expressions, where context plays a crucial role in determining the most appropriate categorisation. As referential expressions are often used to refer to quantities, theories, and findings, their categorisation may depend on whether they are emphasising an aspect or a measurable quantity. Another example of ambiguity arises with the phrase *in the present study*, which may serve as both a location function (indicating the context of the study) and a structuring signal (marking the organisation of the text). As referential expressions are widely used in academic texts to refer to locations, theories, and findings, their function often depends on the specific context in which they are employed, such as in journal articles or textbooks. This illustrates that some lexical bundles can function in multiple ways, depending on how they are used within the discourse.

Secondly, text-oriented bundles function as connective devices. Hyland (2008: 14) defines a text-oriented function as 'the organisation of the text and its meaning as a message or argument'. In his review of text-oriented bundles, he lists four sub-functions that are used to structure academic texts. These are *transition signals, resultative signals, structuring signals* and *framing signals.* A transition signal refers to the 'additive or contrastive links between elements', such as *on the other hand, in addition to the*, *in contrast to* (cf. Hyland, 2008: 14), while the resultative function is to 'mark inferential or causative relations between elements', such as *as a result of, it was found that, these results suggest that* (see also Hyland, 2008:14). Hyland (2008: 14) defines structuring signals as 'text-reflexive markers which organise stretches of discourse or direct the reader elsewhere in text', such as *in the present study, in the next section, as shown in figure*. In this connection, framing signals are used to 'situate arguments by specifying limiting conditions', such as *in the case of, with respect to the, on the basis of* – see Table 5, above.

As can be seen in the above table, the last function is participant-oriented, as it is based on the writer or reader in the text (Hyland, 2008: 14). This function is subdivided into two: stance features and engagement features. Stance features are used by the writer to convey his/her attitudes and evaluations, such as *are likely to be, may be due to, it is possible that*. The writer includes engagement features to involve the reader directly while reading the text, such as *should be noted that, as can be seen*.

Regarding lexical bundles' functions from these two concepts, they have similarities in how they provide common communicative purposes found in academic genres in both spoken and written languages. As such, the functions established by Hyland (2008) are robust regarding the interaction between writers and readers, particularly as found in the academic writing of high-level proficiency students across disciplines. Likewise, Biber, Conrad, & Cortes (2004) propose functions that are normally found in academic registers (cf. Biber et al., 2004; Nesi & Basturkmen, 2006; Mackiewicz, 2016), including functions that are found in a wide range of research articles, particularly in research articles across disciplines (cf. Cortes, 2004, 2006; Hyland, 2008; Csomay, 2013). Since the current research investigates lexical bundles in an engineering research articles (ERAs) corpus, I

adopt the functional taxonomy of Biber, Conrad, & Cortes (2004). More specifically, the functions from Biber, Conrad, & Cortes (2004) suit the current research best, since the specification of attributes, i.e., tangible framing attributes (e.g., *the size of the*) and intangible framing attributes (e.g., *in terms of*), are commonly found in engineering research articles (see Chapter 3). Similarly, it is found that writers frequently use time/place/text references (e.g., *as shown in figure*) to interact with the reader in the text (see also Chapter 3).

Lexical bundles are recurrent sequences of words that frequently occur in specific registers, and they serve as essential building blocks in discourse (Biber & Barbieri, 2007). They are frequently examined for their syntactic and lexical features, which are characterised by their frequency of occurrence. The analysis of lexical bundles is characterised by four main criteria: syntactic fixity, lexical fixity, contiguity, and the transparency or opaqueness of their meanings. These criteria provide insights into the functioning of lexical bundles in a variety of contexts, especially in academic writing and spoken discourse.

The first criterion is syntactic fixity, which refers to the structural patterns of lexical bundles and their embeddedness within discourse. Lexical bundles are frequently recurrent in specific registers and are frequently incomplete structural units. For example, in academic prose, lexical bundles are primarily phrasal, comprising components of noun phrases or prepositional phrases. Conversely, in conversational discourse, they are more prevalently clausal. This distinction was highlighted by Biber et al. (1999), who observed that academic writing is dominated by phrasal bundles, such as those that contain of phrase fragments. This syntactic fixity ensures that lexical bundles serve specific structural functions in discourse, thereby enhancing the overall coherence of communication.

Secondly, lexical fixity, which investigates the frequency and consistency with which lexical bundles appear in discourse. These bundles can be distinguished from idioms or other word combinations by their high frequency of appearance across texts. The determination of their fixedness requires frequent occurrences in numerous contexts, as evidenced by empirical analysis. This frequency-driven identification ensures that lexical

bundles are reliable units for study, as they reflect continuous use of languages patterns. Their lexical fixity makes them predictable, which is particularly important in academic and professional registers, in contrast to flexible expressions.

Third, contiguity is the uninterrupted nature of lexical bundles in discourse. The words in these bundles are continuous sequences that do not span across punctuation or turn boundaries. Their contiguity enables them to function as cohesive units within a sentence, thereby contributing to the perception of fluency in communication. In discourse, for instance, uninterrupted sequences such as "the results of the" retain their unity and meaning. This characteristic ensures that lexical bundles can function as coherence markers, helping the effective organisation and linking of ideas in both written and spoken contexts.

Fourth, the transparency or opaqueness of lexical bundles determines their transparency or opaqueness. Lexical bundles are generally transparent, in contrast to idioms, which frequently have opaque meanings. Their meanings can be derived from the individual words that comprise them, making them accessible to writers and learners. For instance, bundles like *as a result of* or *in the context of* provide clear pragmatic frames for understanding discourse. Their functional roles include expressing stance, organising information, and framing referential content, making them essential tools in both academic writing and spoken communication.

It is evident that analysing of lexical bundles through syntactic fixity, lexical fixity, contiguity, and transparency offers valuable insights into their function in discourse. These criteria highlight the extent to which lexical bundles enhance the coherence, fluency, and clarity of both spoken and written language. Future research could investigate the cognitive processing of lexical bundles and their acquisition in language learning contexts, thereby enhancing our understanding of their communicative functions.

2.4.4 Previous studies on lexical bundles

As previously noted in sections 2.4.2 and 2.4.3, Biber et al.'s (1999) analysis of structural classification and functional taxonomies has contributed to the analysis of lexical

sequences in academic genres, and in pedagogical applications (cf. Cortes, 2002; Chen & Baker, 2010; Gray, 2016). This section aims to include research findings that are related to lexical bundles' structural and functional classification in academic registers, as illustrated in Table 6, below.

Author(s)	Year	Corpus	Corpus Size (words)
Biber, Johansson, Leech, Conrad, & Finegan	1999	Longman Grammar of Spoken and Written English (LSWE Corpus)	40 million
Cortes	2002	Native freshmen compositions	360,704
Cortes	2004	Published writings and student writings	Published writings: 1,992,531 Student writings: 904,376
Biber, Conrad, & Cortes	2004	TOEFL 2000 Spoken and Written Academic Language corpus (T2K-SWAL Corpus)	2,009,400
Biber	2006	TOEFL 2000 Spoken and Written Academic Language corpus (T2K-SWAL Corpus)	2,009,400
Nesi & Basturkmen	2006	Lectures from the British Academic Spoken English (BASE) corpus1 and the Michigan Corpus of Academic Spoken English (MICASE)	1,270,798
Biber & Barbieri	2007	TOEFL 2000 Spoken and Written Academic Language corpus (T2K-SWAL Corpus)	T2K-SWAL: 2,541,795
Cortes	2008	Published history writing in English and Spanish	English: 1,001,012: Spanish: 1,003,264
Hyland	2008	Research articles, PhD dissertations and MA/MSc theses	3,500,000
Kim	2009	Korean Lexical Bundles in Conversation and Academic Texts	The Sejong Corpus: Conversation corpus: 2,604,054 Academic texts corpus: 3,407,020
Chen & Baker	2010	Freiburg-Lancaster-Oslo/Bergen (FLOB) corpus & the British Academic Written English (BAWE) corpus	FLOB corpus: 1 million BAWE: 6.5 million
Cortes	2013	Published Research Articles Corpus (PRAC)	1,002,748
Csomay	2013	University class session portion of the TOEFL 2000 Spoken and Written Academic Language corpus; Michigan Corpus of Academic Spoken English corpus	TOEFL: 2,009,400 MICASE: 1.8 million
Gray	2016	Q+A corpus	400,000

Table 6 Major studies on lexical bundles

As can be seen in Table 6, above, the research findings on lexical bundles are based on a frequency-driven approach (cf. Biber et al., 2003; Biber, 2006; Biber & Barbieri, 2007). In the same vein, studies of lexical bundles are often derived from comparisons of academic registers within or between corpora, ranging from a large corpus (e.g., LSWE Corpus) to a

small corpus. Those corpora are either available corpora or Do-It-Yourself (DIY) ones (cf. Orr, 2006; Reppen, 2010; Nesi, 2013). To better understand the structures and communicative purposes of lexical bundles, Biber et al. (1999) analysed lexical bundles in spoken and written language in the Longman Grammar of Spoken and Written English (LSWE Corpus).

From the frequency and dispersion criteria used in lexical bundles analysis, Biber et al. (1999) regard lexical bundles as recurrent expressions without idiomaticity or structural status. A 40 million-word corpus is used to identify frequent lexical bundles and the cut-off point is set to at least ten times per million words. In terms of the dispersion range, lexical bundles should be distributed across at least five different texts. The bundles generated are interpreted in structural and grammatical terms. Structural interpretation focuses on the organisation of the bundles, such as whether they are main clause fragments, question fragments, or dependent clause fragments. Grammatical interpretation, on the other hand, examines the linguistic forms and syntactic roles within the bundles, such as noun phrases or prepositional phrase fragments. Similarly, lexical bundles normally bridge two different structural units, such as a main clause plus the start of an embedded complement clause, e.g.,*I don't know why*, or a head noun phrase plus the start of an embedded prepositional phrase, e.g.,*the nature of the*. Drawing on the structural and functional taxonomies in Biber et al. (1999), many studies use the structures and functions of lexical bundles to explain the lexical sequences in different registers regarding their frequency and discourse function (see Table 6, above).

Based on corpus findings, lexical bundles have also been classified structurally in academic registers across disciplines. However, the cut-off point (10 times per million word) proposed by Biber et al. (1999) has been adjusted to 20, 40 and 100 per million words, depending to the size and purpose of the research. In Cortes' (2002, 2004) studies, she found that smaller corpora may have proportionally more lexical bundles than larger corpora. Cortes (2002, 2004) describes the characteristics of lexical bundles in research articles in biology. Her study also shows some major structural and a few functional differences between these two disciplines in their uses of bundles.

Regarding academic registers, Biber, Conrad, and Cortes (2004) compare the distribution and functions of lexical bundles in four registers: conversation, university classroom teaching, university textbooks and published academic research writing. They find the lexical bundles are more commonly found in speech than in writing (Cortes, 2015). The structural units of lexical bundles are used differently. For example, there is extensive use of verb and clause fragments in conversation. In contrast, academic writing uses many noun phrases and prepositional phrase fragments. This study classifies bundles by their structural patterns and by a functional taxonomy which includes stance expressions, discourse organizers and referential expressions.

A number of studies have developed functional classifications of these word combinations (e.g., Biber, Conrad & Cortes, 2003; Biber et al., 2004; Biber & Barbieri, 2007). Based on an initial taxonomy of lexical bundles, Biber (2006) adopts the Longman Grammar structures and functions of lexical bundles. The chapter 'Lexical bundles in university teaching and textbooks' opens with a review of studies on multi-word sequences and describes the general characteristics of lexical bundles, which Biber defines as 'the most frequently recurring sequences of words' (p. 133). Lexical bundles in university discourse are viewed as structural 'frames', which allow the interpretation of 'new' information with regard to stance, discourse organization or referential status. Nesi and Basturkmen (2006) investigate the lexical bundles used in monologic lectures. Particularly, they focus on cohesive devices that are used in speech events. Within academic registers, Biber and Barbieri (2007) analyse lexical bundles that are used in spoken and written university registers, including both instructional registers and student advising/management registers (e.g., office hours, class management talk, written syllabi etc.). The findings show that lexical bundles are employed in non-academic university registers less than they are in core instructional registers. Interestingly, lexical bundles are very common in written course management (e.g., course syllabi), in contrast to previous research which found bundles to be much more common in speech than in writing. In the same vein, Csomay (2013) is interested in investigating lexical bundles in spoken lectures. As such, lectures are categorised into discourse units, based on the content words that

occur. The functional taxonomy in Biber et al. (2004) has been used to identify functions that are used in lectures. In particular, stance bundles are normally found in the opening phases of a lecture.

In addition, the use of lexical bundles in academic registers has been paid attention to by researchers in language teaching and learning. Cortes (2004) compares the use of lexical bundles between students and published researchers in history and biology, while Chen and Baker (2010) compare the use of lexical bundles by native and non-native groups.

Hyland (2008) and Cortes (2013) classify the functions of lexical bundles in academic registers. Hyland (2008) selects only 4-word bundles to analyse the functions drawn from Biber et al. (2004). His data were articles and doctoral and masters' theses in the disciplines of biology, engineering, business and applied linguistics. The analysis reveals that there is a high percentage of research-oriented bundles in the engineering discipline, such as the use of location bundles to indicate time and place (*at the same time, at the present study*), procedure (e.g., *the use of, the purpose of*), quantification (e.g., *the wide range of*), description (e.g., *the structure of*, *the size of*) and topic (e.g., *in Hong Kong, the currency board*). Cortes (2013) generates lexical bundles in the published research articles introduction section. She groups texts into moves and steps. Then, lexical bundles are analysed in terms of the internal distribution of moves. From her study, it is suggested that 'some bundles have particular functions tied to specific moves and steps' (Gray & Biber, 2015: 131).

The studies of Csomay (2013) and Cortes (2013) address particular functions of lexical bundles in spoken and written texts. The functions found are also related to the internal organization of discourse. One study that is noteworthy can be seen in Gray (2016), which identifies 4-word bundles from a Q+A corpus. In her study, she does not only employ structure and function in the analysis of lexical bundles, but also uses triangulation methods to determine the reliability of lexical bundles categorisation. She concludes that Q+A is similar to face-to-face spoken registers. The results suggest that there is direct contact between participants in the discourse. The functions found that are relevant to the spoken register are those of epistemic, attitudinal, modal and evaluative stance. In a similar

vein, Gray (2016) applies intra-rater reliability in her work in order to check reliability in lexical bundles' structure and functions' categorisation.

The study of lexical bundles in other languages can be seen in Cortes (2008) and Kim (2009). In a comparative study, Cortes (2008) compares the use of lexical bundles in published history articles in English and Spanish. Later, Cortes compared the lexical bundles identified in those corpora and analysed them in terms of both structure and function. Another recent study exploring a language other than English is Kim (2009). Based on a large Korean corpus, she found that lexical bundles are important expressions in Korean, which function as discourse frames in communication.

As previously mentioned (§ 2.2.1), lexical bundles have been used as one method to analyse and categorise structures and functions in academic registers based on analyses of corpora (notably Biber and his colleagues' work on lexical bundles, see also Biber et al., 1999; 2003; 2004; 2007; Conrad & Biber, 2005; Cortes, 2008; Cortes, 2015; Gray, 2015). Their work is relevant to my study in terms of dealing with lexical bundles' patterns in engineering research articles (see Chapter 3). In section 2.5, the application of corpora in language pedagogy is presented, including the use of phraseology in language teaching.

2.5 Research findings on the application of language corpora and phraseology in language pedagogy

As previously mentioned (§ 2.2), corpus linguistics offers methods for analysing linguistic features in academic genres, such as phraseological patterns (cf. Altenberg, 1998; Gledhill, 2011), collocations & colligations (cf. Renouf & Sinclair, 1991; Sinclair, 1991; Xiao, 2015) and lexical bundles (cf. Biber et al., 1999; Hyland, 2008; Cortes, 2013). In addition, corpora can have an effect on developing learners' language skills (see also Hunston, 2002). Pedagogically, the teaching of language description that is used in general or specific texts may not be well integrated into commercial textbooks or dictionaries (cf. Flowerdew, 2010). Therefore, corpus linguistics provides not only insights into the language use of native or non-native speakers in academic genres but also guidance on which linguistic features should be taught (cf. Fligelstone, 1993; Tognini-Bonelli, 2001;

McEnery & Gabrielatos, 2008). This section makes some remarks about the benefit of using language corpora to teach language skills. Such application reveals not only particular techniques that are used to enhance learners' English skills but also enables learners to develop their English abilities more independently. (cf. Braun, 2005).

Since the focus of this current study is to help learners to write research articles by focusing on lexical bundles, I followed the approach of corpora use in teaching from Leech's (1997: 6) direct use of corpora in teaching, as we can use corpora in 'teaching to exploit corpora so the learners can explore them for their own purposes in various fields'. As such, it focuses on teaching the use of corpora to learners or 'hands-on' know-how, and then learners can explore corpora for their own purposes (Leech, 1997: 8). Similarly, this study is based on the indirect use of corpora for teaching and for developing materials that focus on linguistic features identified in corpora (cf. Leech, 1997; Aston, 2000).

There are four components that are usually integrated into the application of corpora in language teaching: language description, data-driven learning approach, materials design, classroom management and evaluation. As such, this section also focuses on four components commonly highlighted in corpus-based pedagogy studies (e.g., Boulton, 2012; Römer, 2010). These components are: corpus analysis for language description (phraseology), data-driven learning, using corpus-informed approaches to design teaching materials, and classroom management and evaluation.

2.5.1 Corpus-based language description that benefits language teaching

The first component includes analysing of target language features derived from corpus analysis with regard to selected criteria (Cheng, 2010a, 2010b). In this respect, researchers use corpora to help them identify the linguistic features that are used in authentic texts (Braun, 2006). Here, I start by considering the research that is involved in the analysis of a phraseological inventory and which is of benefit to language teaching, especially in the context of this study. It should be noted that lexical bundles are related to the study of lexicology but differ in terms of their classification and discourse functions. While lexicology focuses on the study of words and their meanings, lexical bundles are

sequences of words that frequently occur together and serve specific structural or functional purposes in discourse. The study of lexical bundles typically involves analysing their usage within authentic contexts in corpora, particularly by examining the surrounding words and the patterns in which these bundles co-occur (Prodromou, 2008). The research findings in this aspect concern inventories of formulaic sequences used in academic genres in both spoken and written texts, since phraseological knowledge is one of the key components of learning a language, such as writing (cf. Alderson et al., 1995; Howarth, 1998) and reading (cf. Kremmel et al., 2017). Before writing a complete text, students should first be taught phraseological knowledge, as well as the structural organisation of texts. Regarding the view of corpus linguistics, Römer (2009) states that vocabulary and syntax are inseparable. Howarth (1998) found that one of the factors hindering academic writing is a lack of phraseological awareness of the texts that learners are working on. Consequently, many studies on academic discourse have paid attention to the teaching of phraseological knowledge, such as the structure of lexical bundles and functions in academic discourse (cf.; Biber, 2006; Nesi & Basturkmen, 2006; Hyland, 2008; Cortes, 2008, 2013; see also section 2.4). The structure and functions of lexical sequences are typically examined to understand the meanings they convey within a given context.

Based on studies of formulaic sequences, Nattinger and Decarrico (1992) offer a system for categorizing these phrasal chunks, from fixed items such as proverbs and idioms to syntactic structural templates which can be filled in a variety of ways. Biber et al. (2004) investigated lexical bundles in written classroom teaching and textbooks based on a frequency-driven approach. Similarly, Nesi and Basturkmen (2006) examined 4-word lexical bundles used in monologic university lectures from two different corpora: the British Academic Spoken English (BASE) corpus and the Michigan Corpus of Academic Spoken English (MICASE). In the same vein, Simpson-Vlach & Ellis (2010) invented a list of formulaic sequences called the Academic Formulas List (AFL), which is used in spoken and written language and compared to the Academic Word List compiled by Coxhead (2000). The selection of formulaic sequences is based on frequently recurrent patterns in spoken and written corpora. To include validity in the selection of AFL, they

use ratings of the AFL from experienced instructors to gauge the teachability of the AFL. Similarly, Kazemi et al. (2014) also use rating results from academic instructors to identify lexical bundles that are not necessarily used in research papers in Applied Linguistics. However, they identify lexical bundles that are considered as overused, underused or moderately used, and which have an effect on learners' writing.

Studies of lexical words using small and large corpora are often done in specific academic fields. Mudraya (2006) focused on creating the Student Engineering English Corpus (SEEC), which is based on engineering textbooks to generate lexical items and co-occurrences. The research mostly highlights a lexical methodology for instructing collocations and word combinations. In contrast to Mudraya's emphasis on the wide range of vocabulary used in engineering textbooks, this study distinctly differs and examines the structure and function of lexical bundles. These lexical bundles are used to create materials specifically designed for writing research papers in the engineering field. Consequently, whereas Mudraya (2006) established the foundation for corpus-based language instruction, my study uses corpus findings to meet the requirements of students involved in academic writing and research.

Cheung (2010) explored the Hong Kong Engineering Corpus, focusing on providing learners with the language tools necessary for professional communication in engineering. Cheung's study is part of a broader trend in learner corpus research, which aims to enhance language acquisition through specialised corpora. My study builds on this by specifically focusing on lexical bundles that are most commonly used in engineering research articles. Unlike Cheung's work, which uses a learner corpus to understand language acquisition at a professional level, my study aims to bridge the gap between general engineering language and the specific needs of academic writing, particularly in research articles.

Graham (2014) developed the Engineering Phrases List (EPL) using the analysis of engineering textbooks. His research is on formulaic language, with the objective of identifying prevalent words for instruction in engineering situations. Although Graham (2014) offers significant insights into formulaic expressions within engineering discourse,

my research takes a different approach by not only finding these lexical bundles but also examining their structural and functional significance in academic writing. Furthermore, I emphasise creating practical lessons derived from these lexical bundles to enhance writing proficiency in research papers, an aspect inadequately addressed in Graham's work.

In summary, the primary distinction resides in the emphasis of my study: whereas Mudraya (2006), Cheung (2010), and Graham (2014) make substantial contributions to corpus-based teaching and professional language learning, my research specifically investigates the structure and function of lexical bundles in academic writing and develops materials aimed at enhancing students' research article writing skills.

2.5.2 The Data-Driven Learning (DDL) approach

Along these lines, the concept that is usually employed in teaching corpora to learners is a data-driven learning (DDL) approach. Activities that are based on DDL can encourage learners to consult a corpus according to their needs, and in the end this can increase the degree of autonomous learning (cf. Bernardini, 2002; Braun, 2005).

The concept of a data-driven learning (DDL) approach usually refers to the use of corpora in language teaching and learning which are relevant to teachers and students (cf. Bernardini, 2004; Mukherjee, 2006). This idea of a DDL approach is influenced by John (1991). John and King (1991: iii) define DDL as follows:

> *[DDL is] the use in the classroom of computer-generated concordances to get the students to explore regularities of patterning in the target language, and the development of activities and exercises based on concordance output.*

As can be inferred from this definition, DDL relates to learning a target language as well as setting up language activities based on corpus data. The use of DDL includes interpreting language patterns that are revealed in concordance lines or sentences (see also Hunston, 2002).

In parallel, the concept of DDL, according to Leech (1997: 10), refers to learners as researchers, and it is based on the following viewpoint:

> *The critical and argumentative types of essay assignment […] should be balanced with the type of assignment […] which invites the student to obtain, organise, and study real-language data according to individual choice. This latter type of task gives the students a realistic expectation of breaking new ground as 'researcher', doing something which is a unique and individual contribution, rather than a working evaluation of the research of others*.

In Leech (1997), the differentiation between hard and soft versions of Data-Driven Learning (DDL) is crucial for understanding how to implement corpora effectively in language classrooms. Hard DDL encourages learner autonomy by promoting independent exploration of language data, fostering critical thinking and analytical skills. This approach allows learners to engage directly with authentic texts, helping them identify complex language patterns, such as collocations and syntax, which go beyond traditional learning methods. Learners gain a deeper understanding of language use, which is particularly beneficial for those aiming to master advanced or specialised language areas, such as academic or technical English. Moreover, hard DDL prepares learners for real-world communication, as they analyse data from authentic sources, acquiring practical language skills needed in professional contexts.

However, hard DDL can be challenging and time-consuming, especially for beginners. It requires familiarity with corpus tools and analytical methods, which may overwhelm less experienced learners. The process of interpreting raw data can be complex and time-intensive, making it unsuitable for time-constrained environments. Hard DDL is best suited for learners with a solid grasp of the language and those engaged in research or advanced language studies, where deep analysis and independent learning are necessary.

In contrast, soft DDL offers a more structured approach, with the teacher providing guidance and context. This makes it accessible to learners at various proficiency levels, especially beginners or those who need more support. Soft DDL is flexible, as it can be easily integrated into traditional teaching methods, allowing instructors to combine corpus-based learning with conventional instruction. This approach reduces cognitive load,

making the learning process more manageable for students who may find raw corpus data overwhelming. Additionally, soft DDL is beneficial in time-constrained courses, as it allows instructors to focus on key language patterns without requiring students to analyse extensive data.

However, soft DDL has limitations. It doesn't foster as much independence or critical thinking as hard DDL, as the data is often pre-selected by the teacher. This can limit learners' exposure to a variety of language patterns and may introduce bias based on the teacher's choices. Soft DDL is best suited for beginners or intermediate learners, as it provides a smooth transition to more independent learning.

Although DDL has numerous advantages, some studies critique its widespread use. Dellar (2003), for example, questions the practical impact of corpora on language teaching, suggesting that while DDL has the potential to enhance learning, it may not be as transformative as some proponents claim. Dellar highlights concerns such as the time and effort required to incorporate corpora effectively into the curriculum, especially when teachers are not sufficiently trained in corpus linguistics. These critiques emphasisethe importance of understanding when DDL is appropriate and when other methods might be more effective, particularly for less experienced learners or time-constrained courses.

By incorporating Leech (1997) and Dellar (2003), we can provide a more comprehensive analysis that addresses the strengths and limitations of DDL, guiding educators in choosing the most suitable approach for their teaching context. The choice between hard and soft DDL depends on learners' proficiency, course objectives, and available resources. Hard DDL is ideal for advanced learners or research-based courses, while soft DDL is better suited for beginners or in courses with limited time. Both approaches offer valuable benefits and should be chosen according to the specific needs of the learners and course goals.

As such, the main aim of the DDL approach is to raise students' awareness of the language used in texts (cf. Hunston, 2002; Bernardini, 2004; Mukherjee, 2006).

Although learners are encouraged to take an active role in exploring language use through concordances, teachers still play an important role as facilitators. In this study, the

DDL approach supports student learning by guiding them step-by-step in understanding linguistic features in concordance lines (Bernardini, 2004). This aligns with the Data-Driven Learning (DDL) approach, which emphasizes a collaborative process where both the teacher and the learners play key roles. Thus, at this point, I focus on the concept of DDL for the simple reason that it allows me to introduce the linguistic features used in concordances and can guide engineering students' know-how regarding using a corpus in their learning and at their own pace. However, the degree of help from teachers in the classroom might vary because it depends on the nature of the learners, e.g.,level of proficiency, field, experience of using corpora (cf. Sripichan, 2010; Charles, 2015; Anthony, 2017).

In the above discussion of the DDL approach, we have seen that it provides guidelines for using language corpora in the classroom. However, the DDL approach also has some weaknesses in that there are no suitable corpora for teaching a discipline-specific field to learners (Anthony, 2017). Anthony (2017) explains that the use of DDL is normally based on the teaching of a general corpus to students in different disciplines, and this might not be provided for those who are learning English in a discipline-specific field.

2.5.3 Designing materials based on corpus data

The DDL approach is often integrated with resources that use genuine language data, allowing learners to interact directly with real-world language usage. Examples of such materials include concordance lines, which demonstrate the contextual use of words and phrases to explain language patterns, collocations, and frequency of usage (Ozdemir, 2014). Corpus-based online tools enable learners to interactively examine language data, presenting concordance lines and frequency distributions to study linguistic characteristics, including source use in academic writing (Le, 2017). Furthermore, in-house English for Specific Purposes (ESP) materials, particularly those designed for medical English instruction, use corpus data to develop unique exercises that emphasise collocations relevant to certain professional domains (Ozdemir, 2014). These tools are essential for enhancing language awareness, autonomy, and competency, while offering learners real, contextualised language experience (Boulton, 2017). Following the introduction of corpora

in the classroom by Hunston (2002), McEnery and Wilson (2001) and Sripichan (2010), the processes of using language corpora can be outlined as follows:

- Teachers assess students' knowledge of corpora to determine their familiarity with this resource.
- Teachers identify task objectives and select appropriate corpora, ensuring the materials align with the learning goals.
- Teachers prepare corpus data, tailoring it to match the specific needs of the students.
- Teachers introduce students to corpus analysis tools, such as AntConc, and guide them in exploring language data interactively.
- Both teachers and students interpret the corpus results, with teachers scaffolding the process and helping students understand linguistic patterns and trends.

These five components can be used to prepare students for using corpora. Since students might have different levels of background knowledge of corpora, it is helpful to start by introducing corpora. Then, informing students about the objectives of using corpora in the classroom and types of corpora is also suggested, e.g., introducing the use of the BNC corpus to learn British spoken language. Supporting students in corpus use is the teacher's role, which includes introducing available corpora, helping learners build or compile their own, and guiding them in querying and analysing data effectively. This process ensures learners are equipped to work with both existing and self-compiled datasets. Moreover, it is important to introduce corpus analysis tools and technical support. At this stage, familiarity with concordances is suggested. This involves introducing students to the basic features of concordance tools, such as searching for keywords, analysing word patterns, and observing how words are used in context. The teacher should guide students in interpreting concordance results and identifying language structures and collocations, which are essential for understanding authentic language use. Finally, training learners to interpret corpus data is also important, especially in helping them understand the context and limitations of such data. Teachers should encourage learners to interpret results critically, recognising that frequency counts or patterns may not reflect general language use or be valid across all contexts. To avoid overgeneralisation or

misinterpretation, students need to consider variables such as register, genre, and cultural background when drawing conclusions (see also Ädel, 2010). These processes can be used as a circle in implementing language corpora in the classroom. In addition, classroom management and evaluation of DDL use can also be used in research findings This involves analysing how effectively DDL activities are implemented in the classroom, assessing student engagement, and measuring learning outcomes. Research findings from these evaluations can provide insights into the benefits and challenges of using DDL in teaching. For instance, they can inform the refinement of teaching strategies, the development of more effective materials, and the identification of best practices for integrating DDL into various educational contexts.

2.5.4 Classroom management and evaluation

In terms of classroom management, facilitators can manage lessons while delivering materials in the classroom. This includes organising the sequence of activities; guiding learners in using the DDL tools; and providing support to ensure that students stay engaged and can effectively interact with the materials. Additionally, classroom management involves assessing the effectiveness of DDL implementation by analysing students' progress and outcomes. Research findings from classroom applications of DDL can also inform adjustments to teaching methods and materials to better suit learners' needs.

For example, the use of Presentation (P), Practice (P), Production (P) might be introduced in the classroom (cf. Ur, 1996). After teaching, the lesson can be assessed by the use of research instruments, such as questionnaires, interviews, field notes (cf. Creswell, 1998). However, the use of teaching frameworks, such as PPP (Presentation, Practice, and Production), and evaluation processes depends on the purposes of individual research (cf. Bernardini, 2002; Chamber, 2005; Götz and Mukherjee, 2006). Teaching frameworks like PPP are used to guide the selection and sequencing of teaching materials and procedures, ensuring that lessons are structured effectively to meet learning objectives. Classroom management, on the other hand, focuses on facilitating the implementation of these frameworks, creating an environment where students can actively engage with the materials and achieve the intended outcomes (cf. Bernardini, 2002; Chamber, 2005; Götz and Mukherjee, 2006).

Consequently, DDL has been used to create activities and materials that are specifically tailored to the instruction of writing skills to EAP students. Such activities are typically based on the use of concordances and noticing the use of linguistic features in corpora. For example, students may work on activities and at the same time consult a corpus (cf. Bernardini, 2002; Chamber, 2005). In a similar way, Götz and Mukherjee (2006) evaluate learners' use of DDL activities before and after a 2-hour session. Based on learners' responses in a questionnaire, it reveals that even though DDL activities seem interesting, learners do not learn much from DDL. Questionnaires are useful in the establishment of learning outcomes by capturing learners' perceptions of their progress,

their level of engagement, and their comprehension of the activities. Nevertheless, in order to accurately evaluate the efficacy of DDL activities in enhancing language skills, it is necessary to supplement this self-reported data with performance-based assessments, such as pre- and post-tests. By evaluating the efficacy and challenges of Data-Driven Learning (DDL) in language acquisition, Götz and Mukherjee (2006) assessed the use of DDL by learners. The potential benefits and challenges of using corpora for pedagogical purposes were revealed through the use of questionnaires to investigate how learners engaged with DDL activities.

In this present study, the use of questionnaires in corpus-based language pedagogy provided valuable evidence for investigating learning outcomes and student engagement with DDL activities. Questionnaires proved effective in evaluating students' perceptions, measuring confidence levels, and assessing their understanding of lexical bundles. For instance, the questionnaires evaluated the frequency with which students consulted corpus data and their perceived improvements. However, several challenges arose regarding reliability and validity. Factors such as self-perception bias and the clarity of questions may have affected the consistency of responses (Eckerdal & Hagström, 2017). Pilot testing helped to mitigate these issues by refining the questionnaire design, while combining questionnaire findings with other assessment methods, such as analysing students' performance using corpus tools, ensured validity. Moreover, the inclusion of open-ended questions fostered deeper reflection and enhanced engagement, as students articulated their learning processes and challenges. Overall, in this present study, questionnaires, when implemented thoughtfully, complemented other assessment tools to provide comprehensive evidence of students' exploration of language structures and functions. Their results are consistent with the broader body of research on DDL, which emphasises its importance in supplying authentic language data for learning while also recognising the necessity of thorough implementation and support to overcome its challenges.

The incorporation of DDL elements in ESP materials for teaching English in discipline-specific fields is exemplified in the works of Charles (2015) and Anthony (2017). These studies highlight how corpora can be effectively integrated with materials

development to address the specific linguistic needs of learners in specialised contexts, such as academic and professional fields. Charles has incorporated corpus use in EAP classes. She included materials developed by analysing the linguistic features found in a learners' writing corpus at postgraduate level. In her study, the learners were required to use corpus tools, specifically AntConc, to analyse their writing (Charle, 2015).

Additionally, the tools consisted of wordlists, which offer a comprehensive list of all words in the corpus along with their frequencies. Concordance lines were also used to analyse the usage of search terms, clusters were implemented to investigate groups of words, collocates were implemented to identify words that frequently co-occurred, keyword lists were implemented to identify words that were unusually frequent or infrequent, and n-grams were implemented to analyse groups of words of a specific size.

Learners were able to identify and evaluate the use of lexical words in their writing with the aid of these tools. She found that her work could engage students in working on concordances; however, a weakness was also found. Charles (2015) explains that since learners are from mixed disciplines, their interpretations of concordances are contradictory. This is a result of the fact that learners' disciplinary background have an impact on their ability to understand and analyse linguistic patterns. For instance, students in technical fields may prioritise precise terminology, whereas those in the humanities may interpret language more subjectively, resulting in varying conclusions. Anthony (2017) shares similar findings, since it is suggested that a corpus should be more specific to an individual group of students.

Cortes (2006) has applied lexical bundles in teaching history to native speaker students. She developed materials that included the structure and functions of lexical bundles, demonstrating that their use can enhance students' awareness of how to use expressions effectively. In particular, Cortes (2006) indicates that there is no significant difference in student learning outcomes between the pre-test and post-test following instruction that incorporated lexical bundles. The results from Cortes (2006) can be explained thus: the students are native speakers of English so they have some background knowledge in the use of lexical bundles in the corpus. Conrad (2017) focuses on the

investigation of passives and an impersonal style in student reports and professional writing. Conrad's (2017) study provides insights into the gap between academic writing practices and the writing demands of the engineering profession. It also contributes to the development of instructional materials aimed at better preparing engineering students for the professional writing tasks they will encounter in the workplace.

I have integrated specific corpora (ERAs) into the teaching of lexical bundles (see Chapter 3 for a more detailed discussion of lexical bundles analysis) to engineering students (see Chapter 4 for a more detailed discussion of the procedures for using corpora in the teaching of lexical bundles). I close this section with the following quote on why the application of corpora suits the current investigation:

> *The advances in the direct access to corpora by language teachers and learners have created the need to research into a number of pedagogic issues, including the types of corpora to be consulted, large or small, general or domain-specific, tagged or untagged; the kinds of learning strategies to benefit from direct corpus consultation; and the means by which direct access to corpora can be integrated into the language learning context.*

(see Chambers 2005: 111; emphasis added)

Thus, this study focuses on examining the structure and function of lexical bundles in ERAs and on exploring how corpora can be used as a teaching tool for engineering students. The linguistic aspect centres on analysing and interpreting lexical bundles, while the methodological approach involves applying corpus-based techniques to enhance teaching practices. As such, the application of these two approaches can facilitate students' learning of lexical bundles found in a specific corpus systematically. To overcome the weaknesses of DDL (cf. Anthony, 2017), this current study builds a corpus that is specific to the context of the learners. As such, I have compiled a corpus that is suitable for the students' context of learning, i.e., the ERAs corpus (see Chapter 3). Regarding the use of the DDL approach, the use of concordances as data can encourage learners to explore the use of lexical bundles in the ERAs corpus by themselves (see Chapter 4). This is also in line with the studies of Braun (2006), Braun et al. (2006) and Mukherjee (2006). Therefore, learners' engagement with the process of interpretation of lexical bundles' use in a corpus

is considered the ultimate goal in autonomous learning. This is also supported by Leech (1997), in that when learners engage in the interpretation process themselves, they create models of their own learning.

2.6 Summary

In this chapter, I have focused on four important themes. First, I introduced definitions, concepts and methods that are used in corpus linguistics: corpus-based approach and corpus-driven approach. After that, I referred to previous research on using corpus techniques in analysing academic corpora, i.e., academic wordlists and multi-word expressions (§ 2.2). Later, I introduced the concepts of phraseology that have influenced the two main approaches: Russian phraseology and the Firthian tradition (§ 2.3). Those two approaches have influenced how lexical sequences are differently defined and identified with the use of a corpus: methods used for identification (frequency and dispersion range), types of expressions (continuous, discontinuous, 2-word collocations or longer word combinations, idioms etc.). In section 2.4, I referred to the concept of lexical bundles as one of the constructs in the phraseological spectrum, while specifically highlighting the structure and the functions of lexical bundles. Previous research on lexical bundles is also explained (§ 2.2). Finally, the chapter discusses research findings on the application of language corpora and phraseology in language pedagogy (§ 2.5). The next chapter highlights the data, methods and analyses of general lexical bundles and specific lexical bundles in Engineering Research Articles (ERAs).

3 Chapter Three: Data, methods and analyses of general lexical bundles and specific lexical bundles in Engineering Research Articles (ERAs)

3.1 Introduction

This chapter identifies the characteristics of lexical bundles (LBs) in Engineering Research Articles (ERAs): structure, function, frequency, dispersion range. It outlines methodological applications and issues in lexical bundles analysis (§3.2)**.** First, I present the reasons for choosing engineering research articles (§3.2.1). After that, issues regarding lexical bundles' selection, an inter-rater reliability process and the categorisation of lexical bundles' structure and functions are covered**.** In order to provide a comprehensive analysis, this study includes not only the Top 50 three-word and four-word lexical bundles but also the findings for the Bottom 50 three-word and four-word lexical bundles, which are further discussed in Sections §3.3 and §3.4. The ERAs corpus comprises 100 engineering research articles. The lexical bundles identified from the corpus may be beneficial for engineering students in improving their research writing (§3.4). Section 3.5 explores the overall patterns of the top 50 and bottom 50 lexical bundles, highlighting differences in their categorisation. It also discusses how this study relates to previous research on lexical bundles in academic writing. Finally, I provide a summary of the lexical bundles analytical frameworks employed in my analysis (§3.6).

3.2 Methodology

3.2.1 Corpus used in this study

The corpus I compiled consists of 100 engineering research articles (ERAs). In order to answer my first research question: *What are the generic types of lexical bundles (Top 50), and what are the specific lexical bundles that are more relevant to phrasal expressions in the engineering domain (Bottom 50), that are used in engineering research articles?* I will discuss lexical bundles that are general and specific in ERAs in Sections 3.3 and 3.4 of this thesis. In response to my second research question: *How these generic and specific lexical bundles are used in engineering research articles, in terms of frequency, dispersion range, structure and discourse function,* I will discuss the

characteristics of general and specific lexical bundles in the ERAs corpus regarding the criteria of frequency, dispersion range, structure and discourse function (§3.3 & 3.4).

I chose to study ERAs because they are used in learning and teaching for engineering students at KU.CSC. The corpus was compiled from 100 engineering research articles in leading journals[11] published by Science Direct and issued between 2003 and 2015. The journals[12] included in the data set are as follows: Cement and Concrete Research, Computer and Electrical Engineering, Electrical Power Systems Research and Applied Thermal Engineering (see Appendix B). Each ERA selected is approximately 4,000 words in length and their format is generally: introduction (or background, related work, related theory), methods (or experimental details, modelling, implementations, procedures, problem formulations), results (or analysis) and discussion, and a conclusion (plus acknowledgements and references). The 100 electronic research articles were downloaded in pdf files and then they were converted into a text file. After electronic copies of all the articles were collected, non-textual annotations were erased, e.g., titles, page numbers, tables, statistical graphics, numerical data, formulations and references.

The issue of selecting journals written by native or non-native speakers of English is not the main focus of this study. If their articles are accepted to be published, it means that their writing is of an acceptable standard, since there are strong gatekeepers who screen research articles. The research articles selected from leading journals are considered to represent the 'norm' and have a 'high impact factor',[13] as well as being 'strong gatekeepers'. It is suggested, therefore, that the written English used in these journals is by definition considered to be of an academically acceptable standard, and so it is appropriate

[11] The research articles and journals chosen were based on research articles that the students were using in their learning and teaching as recommended by engineering course instructors. The research articles are chosen from four engineering fields taught at KU.CSC: Power, Electrical and Computer, Civil, Mechanical.

[12] It should be noted that some journals, such as Cement and Concrete Research, cover more than one subdiscipline.

[13] The impact factors of the four journals are 3.480 for Cement and Concrete Research, 1.084 for Computers & Electrical Engineering, 1.809 for Electric Power Systems Research, 3.043 for Applied Thermal Engineering.

for students to learn units of meaning of LBs in engineering research articles from these selected articles.

In terms of generalisability, the 100 ERAs (engineering research articles) were chosen based on their frequent use in the instruction of engineering students at KU.CSC. A random sampling procedure was used to select these publications from a 12-year period (2003-2015) to ensure that the sample reflects the most recent engineering knowledge at the time. However, it is crucial to acknowledge that the potential for generalisation of findings from a small number of journals may be limited. Civil engineering, for instance, contains a variety of subfields, including hydrology, geotechnology, water, and building. Each of these subfields has produced a wide range of journals. Consequently, it is possible that the four journals that were chosen do not reflect all subfields within civil engineering.

The 100 ERAs were randomly selected from ScienceDirect's open-access journals (eg for the cement and concrete journals, https://www.sciencedirect.com/journal/cement-and-concrete-research/issues). The selection of 25 articles from each journal was determined by the available articles within the 12-year period and the relevance of these articles to the teaching and learning context at KU.CSC. The selection of 25 articles from each journal was designed to ensure that the sample was representative of the available articles and that it reflected a variety of topics within the journals, while also maintaining the randomness of the selection process for the sample. In spite of the restricted selection of only four journals, this sample is still relevant to the scope of this investigation (Mackey & Gass, 2015: 172).

With the aid of AntConc (version 3.5.7) a freeware corpus analysis toolkit, created by Laurence Anthony (Anthony, 2011), I could generate lexical bundle lists and utilise concordance lines to determine the functions of target bundles. The selected 100 ERAs were stored electronically, converted into plain text files and then processed with AntConc. The lexical bundles generated show the frequency with which each bundle occurs and the number of research articles in which it was used (distribution range). I set criteria for identifying lexical bundles that are generic and specific to engineering, as outlined in the following section.

3.2.2 Issues related to noise in lexical bundle selection

The main criterion in selecting lexical bundles is to choose uninterrupted ones, meaning those that consist of continuous 3-word and 4-word sequences without any interruptions or breaks between the words. For example, sequences like *in the case of* or *on the other hand* are uninterrupted bundles.

The non-textual annotations were erased. These annotations refer to elements within the corpus that are not part of the main textual data but may still appear in the content. Examples include mathematical formulae, alphabetic characters, symbols, interrupted chemical names, numbers referring to temperature, and citations such as names of authors and journals. This technical problem is found to be in line with the study by McEnery and Baker (2015: 248), whose corpus contained some content words and also symbols or numbers which can generate 'noise', and these are considered to be irrelevant lexical bundles. Since AntConc does not consider numbers and mathematical symbols to be parts of words, some bundles in the findings were difficult to understand and could not be interpreted. Therefore, one of the tasks when processing the data was to get rid of interrupted words and exclude lexical bundles that were acronyms.[14]

Abstracts, acknowledgements, and references were excluded and manually removed at the beginning of the electronic text collection process. Furthermore, page numbers, authors' names, headers, footers, figures, tables, chemical name indicators, and mathematical formulae were manually eliminated to ensure the corpus focused solely on meaningful lexical patterns in research articles. This study focusses on the main sections of research articles, as they most accurately represent the language used in academic research writing in the field of engineering. Abstracts were excluded due to their classification as a distinct genre of academic writing, characterised by unique features and purposes, including a summary of key research points. This study aims to enhance students' understanding of the language used in research article sections, such as the introduction,

[14] In this analysis, acronyms are included, because they refer to specific subjects in the engineering field; they are treated as one sequence of words. For Example, XRD is from X-ray Diffraction and is among NP fragments in corpus identification.

methodology, and discussion. Consequently, abstracts were excluded to ensure alignment with the research objectives. Page numbers, authors' names, headers and footers, figures, tables, characters indicating chemical names and mathematical formulae were manually removed. At times, the lexical bundles in concordances contained unidentified names of chemical substances, so the researcher had to go back and check manually. For example, there are various items of noise that appear in the ERAs corpus, such as *C\xD0S\xD0H* (referring to calcium silicate hydrate), *at 60 \xA1C* (referring to at 60 °C), *xa c in* (referring to Celsius), *et al. computers and electrical engineering*, *ef\xDEciency of* (referring to efficiency of), *the in xdfuence* (referring to the influence), *energy ef xdeciency* (energy efficiency), *de xdened as the ratio of* (referring to defined as the ratio of). Also, noise sometimes caused AntConc to treat 2-word bundles as three-word ones, e.g.,*was de xdened* (was defined), *satis xdees* the (satisfies the), *signi xdecant effect* (significant effect). Three-word bundles may also be treated as 4-word bundles, e.g.,*ef\xDEciency of the* (efficiency of the), *the in xdfuence of* (the influence of), *C\xD0S\xD0H)* (C–S–H). To clean the noise, I have reviewed those bundles that appear in the original files and fixed the noise in continuous sequences. For example, the noise *ef\xDEciency of the* originally came from the bundle *efficiency of the*. I have corrected and selected those bundles that are in 3-word and 4-word sequences. The percentage of the noise is approximately 5% of the generated bundles, and it does not affect the final results of the 200 lexical bundles.

3.2.3 Categorisation of lexical bundles: frequency and dispersion range

Before I go on to explain this criterion, I would refer the reader back to the definition of lexical bundles, as explained in Chapter 2, where lexical bundles are considered to be 'recurrent expressions, regardless of their idiomaticity, and regardless of their structural status' (Biber et al., 199: 990). As discussed there, lexical bundles frequently recur in natural discourse, they comprise continuous strings of words without any empty slots (cf. Cortes, 2015). Therefore, the frequency and dispersion range should be identified before categorising lexical bundles' structure and functions (cf. Chen & Baker, 2010).

This study focuses on continuous sequences of 3-word bundles and 4-word lexical bundles from the Top 50 and Bottom 50 bundles, because the 3-word bundles and 4-word lexical bundles are the most frequent bundles and they show clearer structures and functions than 5-word bundles and 6-word bundles (Biber et al., 2004). Based on the lexical bundles generated from AntConc software, there were 383,605 3-word bundles and 469,823 4-word bundles in total. The Top 50 refers to the 50 most frequent bundles, and they were selected from the list of 3-word bundles. Similarly, the Top 50 for 4-word bundles consists of the 50 most frequent bundles. These selections together make up a total of 100 bundles. The Bottom 50 refers to the 50 least frequent bundles. Fifty 3-word bundles and fifty 4-word bundles from the bottom of the list were selected, totalling 100 bundles. The overall number of bundles in the study, including both the Top 50 and Bottom 50, was 200 bundles in total.

In the same vein, 3-word bundles can be extended to 4-word bundles by inserting function words such as *a, the, of, as*. For example, the bundle *the effect of (139)* is categorised under the structure of *noun phrase based (NP-based)* and it can be extended to 4-word bundle *the effect of the (45)*. Thus, the bundle *the effect of the (45)* belongs to the structure NP with post-modifier fragment. However, 2-word lexical bundles are excluded because of their lack of meaning/ functionality, such as *of the, with the, that the* (cf. Biber et al., 2004, Gray, 2016).

In terms of dispersion range, the top 50 three-word and four-word lexical bundles show frequencies ranging from 400[15] to 24.[16] The dispersion range is between 82[17] and 13[18] out of 100 engineering research articles. The bottom 50 three-word and four-word lexical bundles all had a frequency of 2. The dispersion range was from 2 to 13 out of 100. Then, a refined list of bundles was identified in relation to the structures they have and the functions they perform in engineering research articles. Two hundred lexical bundles from

[15] A maximum frequency of 400 means that the bundles occur 400 times in the ERAs corpus (400,000 words).

[16] A minimum frequency of 24 means that the bundles occur 24 times in the ERAs corpus (400,000 words).

[17] A maximum dispersion range of 82 means the bundles occur in 82 out of 100 ERAs.

[18] A minimum dispersion range of 13 means the bundles occur in 13 out of 100 ERAs.

the top 50 and bottom 50 three-word and four-word lexical bundles were categorised structurally and functionally according to the framework of Biber, Conrad, and Cortes (2004),[19] as illustrated in Table 7. To deal with lexical bundles that are not explicitly labelled, comparisons were made with similar bundles, along with an examination of the concordances of the bundles used in the ERAs corpus (cf. Gray, 2016). As such, this was to verify the most appropriate categories based on the typical use of bundles in the ERAs corpus.

To ensure the accuracy and reliability of the findings, the 200 lexical bundles were classified by two independent raters based on their structure and function. In terms of the suitability of the raters, the two raters were a doctoral student and a doctoral researcher in Applied Linguistics, making them well-suited for ensuring inter-rater reliability. Both raters have a good understanding of the structure of lexical bundles and have extensive experience teaching English to EFL/ESL students. To classify lexical bundles correctly, it is important to fully understand their structure and function. The raters are skilled at identifying and classifying bundles into structural categories, such as clauses, prepositional phrases, verb phrases, and noun phrases. Their knowledge of lexical bundles helps them carefully decide if each bundle belongs to a specific structural category. This ensures the analysis is accurate and consistent. The raters' understanding of bundle features is important because it directly affects their ability to provide reliable and valid evaluations of the lexical bundles in this study.

The raters were trained in the procedure of identifying lexical bundles' structural and functional categorisation. Then, the two raters compared their results of the lexical bundles' structure and function. The researcher coded the agreements and disagreements between both raters. In cases where lexical bundles were listed without a clear labelling of their structure and functions, final decisions were made through negotiation and by examining all the lexical bundles used in the ERAs corpus. The agreement on lexical

[19] The framework in Biber, Conrad, and Cortes (2004) contains multiple subcategories within each structural and functional type of bundle. Only the main category is used in this study, although it should be noted that most of the subcategories within this taxonomy are represented by the bundles identified in the ERAs corpus.

bundles' structural and functional categorisation between the raters was 97% for each property (see Appendix 4), indicating a high level of inter-rater reliability. I assessed inter-rater reliability using Cohen's kappa (κ). The two raters achieved 97% observed agreement in the categorisation of lexical bundles' structure and function (see Appendix 4), with κ = 0.90, indicating almost perfect agreement according to Landis and Koch's (1977) benchmarks. This demonstrates that the agreement is highly unlikely to be due to chance and supports the reliability of the findings.

The resulting 200 lexical bundles' structural and functional categorisation are listed in frequency order and dispersion range order (see Appendix 3). Sections 3.3 and 3.4 present the distribution of these 200 bundles from the Top 50 and Bottom 50 three-word and four-word lexical bundles by structural type and discourse functional type in the ERAs corpus.

3.3 General lexical bundles in engineering research articles

This section focuses on the ways in which I categorise lexical bundles' structure and functions as they are found in the ERAs corpus (§ 3.3.1 and 3.3.2). I refer to these 100 lexical bundles, consisting of the top 50 three-word and top 50 four-word bundles, as the top 50. They are identified based on their frequency and dispersion range, as specified in terms of frequency and dispersion range (see Appendix 3). I specifically examine the structure of these lexical bundles by following the structural framework from Biber et al. (2004). I also analyse the functions of lexical bundles using Biber et al.'s (2004) functional taxonomies and examine how they perform in the ERAs corpus.

3.3.1 Structures of general lexical bundles in the ERAs corpus

Table 7 shows the distribution of the Top 50 3-word and 4-word lexical bundles by structure. Drawing on the structure framework from Biber et al. (1999), the main structures of lexical bundles from the Top 50 3-word and 4-word lexical bundles are identified. These structures are included in the analysis, i.e., noun phrase fragments, prepositional phrase fragments, passive verb phrases, copula be + adjective phrases, adverbial clause fragments,

anticipatory it + verb/ adjective phrases, verb phrase + that clause fragments, to clause fragments, other expressions and other phrase fragments**.** Following Biber, Conrad, & Cortes (2004), those structures were grouped into four main structural types of lexical bundles**:** noun phrase and prepositional phrase fragments, verb phrase fragments, clausal fragments, and others, as illustrated in Table 7 below**.**

	Structural characteristics of LBS					
Number of LBs	**Preposition-phrase based**	**Noun-phrase based**	**Verb-phrase fragments**	**Clausal fragments**	**Others**	**Total**
Number of 3-word bundle (Percentage)	15(30%)	17(34%)	12(24%)	3(6%)	3(6%)	50(100%)
Number of 4-word bundle (Percentage)	19(38%)	11(22%)	14(28%)	5(10%)	1(2%)	50(100%)
Total	34(34%)	28(28%)	26(26%)	8(10%)	4(4%)	100(100%)

Table 7 Distribution of Top 50 3-word and 4-word lexical bundle types by structure

Table 7 shows the distribution of the Top 50 3-word and 4-word lexical bundle types by structure. The findings indicate that the three most frequent structural types are prepositional phrase fragments (34 out of 100), noun phrases (28 out of 100), and verb phrase fragments (26 out of 100). In contrast, clausal fragments (8 out of 100) and other expressions (4 out of 100) appear less often in the ERAs corpus.

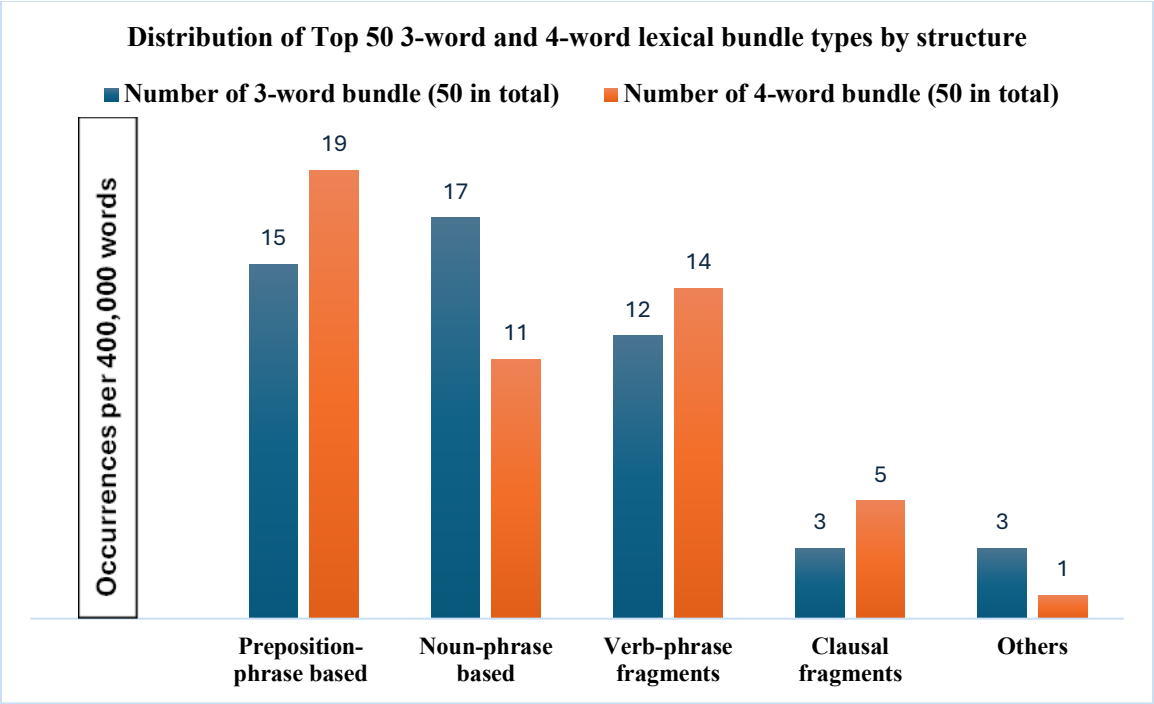


Figure 1 Distribution of Top 50 3-word and 4-word lexical bundle types by structure

Figure 1 also shows the structural distribution of the top 50 three-word and four-word lexical bundles. Prepositional phrase-based and noun phrase-based bundles are the most frequent, with 34 and 28 instances respectively. These are followed by verb phrase-based bundles (26 out of 100). Clausal fragments (8 out of 100) and other types (4 out of 100) are less frequently found in the ERAs corpus.

Structural types of LBs in ERAs corpus	Numbers of LBs (Percentage)	Description	Example
A. Phrasal	34(38%)	Incorporate preposition plus a noun-phrase fragment and another prepositional phrase fragment	
1. Preposition phrase based			
1a. PP with embedded of-phrase fragment			*in terms of (129), in the case of (61)*
1b. Other prepositional phrase (fragment)			*in order to (358), on the other hand (86)*
2. Noun phrase based	28(28%)	Consist of noun phrases with post-modifier fragments	
2a. (connector+) noun phrase with of-phrase fragment			*the number of (202), the effect of the (45)*
2b. Noun phrase with other post modifier fragment			*an increase in the (23)*
3. Verb Phrase based	26(26%)	Incorporate fragments of verb phrase, including subject pronouns followed by a verb phrase, the beginning of a verb phrase	
3a. Anticipatory it+verb phrase/adjective phrase+complement clause			*it can be (171), it can be seen (57)*
3b. Passive verb + prepositional phrase fragment			*shown in fig (400), is shown in fig (114)*
3c. Copula be + noun phrase/adjective phrase			*is due to the (26)*
B. Clausal		Include both verb phrase fragments and components of dependent clauses (e.g., complement clauses)	
1b. (verb/adjective+) to-clause fragment	8(10%)		*can be used (113), can be used to (44)*
2b. (verb phrase+) that-clause fragment			*can be seen (99), can be seen that (31)*
3b. Adverbial clause fragment			*to determine the (73), and can be concluded that (23)*
C. Others	4(4%)	Bundles with structures that do not fit the three main types, including specialised expressions commonly used in academic and technical writing, particularly in fields like engineering.	*the heat transfer (81), as well as the (58)*
Total	**100(100%)**		

Table 8 Structural framework for categorisation of Top 50 3-word and 4-word bundle in ERAs Corpus (adapted from Biber, Conrad, & Cortes 2004)

As illustrated in Table 8, lexical bundles that incorporate noun phrase and prepositional phrase fragments (62 out of 100) often end with the beginning of a postmodifier, such as *in terms of (129)* and *the number of (202).* Verb phrase fragments (26) may either begin with a subject pronoun followed by a verb phrase (e.g., *it can be* (171)), or consist of verb phrase fragments (e.g., *shown in fig.* (400), *is shown in fig.* (114)).

3.3.2 Functions of general lexical bundles in ERAs corpus

As can be seen from the distribution of bundles in the Top 50 in Table 9 below, there are four primary functions served by lexical bundles in the ERAs corpus: referential (50 out of 100), discourse organisers (35 out of 100), stance (14 out of 100) and others (1 out of 100). Drawing on Biber et al**. (**2004**:** 384**)**, referential expressions are used to **'**make direct reference to physical or abstract entities or to textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important**'.** Discourse organisers are used as connective devices to connect texts; Biber et al**. (**2004**:** 384**)** state that discourse organisers **'**reflect the relationships between prior and coming discourse**'**, while stance bundles relate to writers**'** attitudes and assessments **(**cf**.** Biber et al., 2004; Hyland, 2008**).** Other expressions in the findings concern content-based expressions that are specific to the field of engineering, e.g., *the heat transfer (81*), as shown in Table 9**.**

Number of LBs	Discourse functions of LB				
	Referential	Discourse Organisers	Stance	Others	Total
Number of 3-word bundle (Percentage)	25(50%)	21(42%)	3(6%)	1(2%)	50(100%)
Number of 4-word bundle (Percentage)	25(50%)	14(28%)	11(22%)	0(0%)	50(100%)
Total	50(100%)	35(35%)	14(14%)	1(1%)	100(100%)

Table 9 Distribution of Top 50 3-word and 4-word lexical bundle types by discourse function

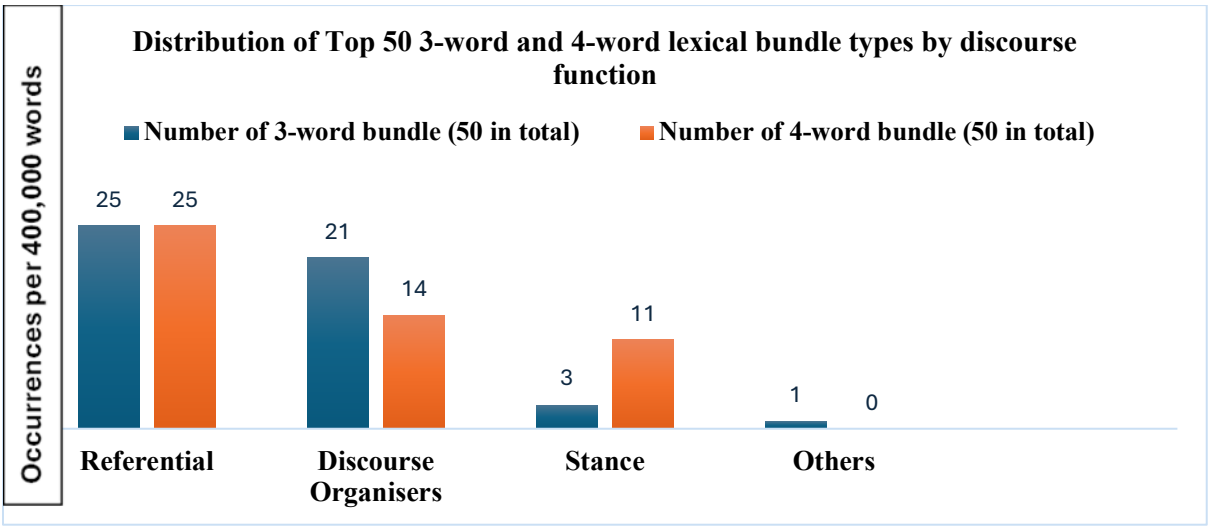


Figure 2 Distribution of 100 Top 50 3-word and 4-word lexical bundles: discourse functions

3.3.2.1 Referential bundles

A preliminary interpretation of these findings is based on the context of the ERAs corpus. The structural trends in Figure 1 above are reflected in the patterns in the most frequent discourse functions for these lexical bundles in the ERAs corpus. Figure 2 shows a consistent trend for conveying referentiality (50%), which is by far the most frequent function of lexical bundles in the ERAs corpus, particularly through the use of noun phrase

based and prepositional phrase based bundles. The findings showing that *referentiality* constitutes 50% of the most frequent functions of lexical bundles in the ERAs corpus are supported by consistent patterns identified across the dataset. This is reflected in the prominent use of noun phrase-based and prepositional phrase-based bundles. Although statistical tests were not explicitly conducted, the consistency and frequency observed across the corpus strongly suggest that these trends are not random but indicative of a significant functional preference in engineering research articles. The analysis of the ERAs corpus includes four main sub-categories of referential bundles: identification/focus, specification of attributes, and time/place/text reference.

A. Identification/focus

Based on the findings from the ERAs corpus in Table 10 below, the use of identification/ focus is commonly found in the ERAs corpus. Identification/focus in this context refers to the emphasis placed on certain elements within a sentence or discourse. Identification involves distinguishing between information focus and identificational focus, which influences how elements are perceived and understood in communication.

Category	Sub-category	Example bundles from ERAs corpus
1. Referential expressions (50%)	A) Identification/Focus	*one of the (100)*
	B) Specific Attributes	
	B. 1) Quantifying specification	*the number of (202), in the range of (28)*
	B.2) Tangible framing attributes	*the value of (81), the temperature of the (34)*
	B.3) Intangible framing attributes	*the use of (161), in the case of (61)*
	C. Time/Place/Text references	
	C. 1) Place reference	*of the system (144)*
	C. 2) Time reference	at the same (81)*,[1] at the same time (56)*
	C. 3) Text deixis	*shown in fig. (400), as shown in fig. (141)*
	C. 4) Multi-functional reference	*the beginning of the (29)*, the end of the (29)**
2. Discourse organisers (35%)	A. Topic Introduction/Focus	*according to the (136), with respect to the (47)*
	B. Topic Elaboration/Clarification	*in order to (358), on the other hand (86)*
3. Stance expressions (14%)	A. Epistemic stance	
	A. 1) Impersonal	*to the fact that (24)*
	B. Attitudinal/Modality stance	
	B. 1) Obligation/directive stance: Impersonal	*it is necessary to (30), it should be noted (29)*
	B. 2) Ability: Impersonal	*it can be (171), it is possible to (44)*
4. Others (1%)		*the heat transfer (81)*

Table 10 Functional type of Top 50 3-word and 4-word lexical bundles in the ERAs Corpus

Regarding the structure of lexical bundles, the NP- based bundle type is included in this subcategory in order to emphasise the main point, as seen in extract (1) below. The bundle *one of the (100)* is used to summarise the main points of methods in ERAs the researcher conducted.

(1) Thus, *one of the* main advantages of our method is that it reduces the high dimensionality to eight DOFs, optimises the number of particles, and achieves real-time hand tracking and high accuracy of 3D full DOF hand gestures by examining CBM for users.

B. Specific attribute bundles

This second subcategory of referential bundles is divided into three main aspects: quantifying specification, tangible framing attributes, and intangible framing attributes. Quantifying specifications focus on measurable elements, such as numbers or ranges. For example, phrases like *the number of* (202 occurrences) and *in the range of* (28 occurrences) highlight numerical or statistical data in engineering contexts. Tangible framing attributes, on the other hand, describe physical or observable features. Examples include *the value of* (81 occurrences) and *the temperature of* (34 occurrences), which provide specific details about systems or materials. Both types are important in academic writing, with one focusing on measurements and the other on how they are described in context.

The purpose of this subcategory is to identify the attributes of a noun. As can be seen in extract (2) below, the bundle *the number of* (202) functions to quantify and specify the amount associated with the head noun alternative paths. This usage highlights its role in describing measurable aspects within the context.

(2) This is because *the number of* alternative paths that a header can take to progress changes as it advances towards the destination.

While the tangible framing attributes bundle *the temperature of (34)* in extract (3) indicates the concrete characteristics of the following noun (i.e.,a fluid), intangible framing attributes bundles help to structure the activities of experiences in the real world (see also Hyland,

2008). In extract (4), the bundle *the use of (161)* helps the writer to structure the activities of engineering experiments in the ERAs corpus (cf. Hyland, 2008).

(3) The throat is located at 59 mm distance in front of the turbine nozzle exit to measure *the temperature of* the fluid leaving the turbine**.**

(4) *The use of* a stationary supercapacitor energy storage device and the reconfiguration of the power system was compared**.**

C. Place/time/text deixis/multi-functional reference bundles

The last subcategory of referential bundles refers to particular place, times or locations in ERAs. In extract (5), the bundle *of the system (144)* expresses belonging. In extract (6), the bundle *at the same (81)* is used as a prepositional device to refer to two properties being at the same amount of heat. It is important to acknowledge that the bundle *at the same* can perform multiple functions simultaneously, including indicating time or providing a reference in a text. For instance, the bundle *at the same time* may serve as a reference in a text or indicate a moment of time.

(5) Security Requirements are defined as constraints on the functions *of the system*, and these constraints operationalise one or more security goals.

(6) As a consequence, the comparison of one property *at the same* amount of heat is not the same as comparing it at

Unsurprisingly, there is extensive use of text deixis bundles in academic written genres, in particular in research articles. Based on the findings from the ERAs corpus, the bundle *as shown in Fig (141)* makes direct reference to figures contained in ERAs in order to present information to the reader, see extract (7) below.

(7) *As shown in Fig.* 5, although the quantity of aluminate in FA is higher than that in GGBS, the quantity of FS in sample CF is still lower than that in sample CG after standard curing for 56 days.

(8) This represents the fourth characteristic point CP4 and *the beginning of the* final stage on the TG curves.

Interestingly, many of these bundles in the ERAs corpus are multi-functional, referring to place and time. For example, the bundle *the beginning of the (29)* in extract (8) can refer to place and time, it depends on the contexts in which bundles are used. Similarly, the bundle *as shown in Fig (141)* mentioned in extract (7) can be used as a marker to organise or structure texts (cf. Hyland, 2008).

3.3.2.2 Discourse organising bundles

The results from the ERAs corpus reveal that there are two main aspects of how discourse organising bundles are used in the ERAs corpus: topic introduction/focus and topic elaboration/clarification (cf. Biber et al., 2004). There is also use of NP-based bundles in discourse organisers.

A. Topic Introduction/focus

As can be seen in extract (9) below, the bundle *according to the (136)* is used to signal to the reader a new topic in ERAs. Similarly, in extract (10) the same expression indicates the focus of what is being introduced.

(9) *According to the* first law of thermodynamics and the Fourier law, unsteady temperature field and heat balance equation in polar coordinates of stator screw drill bushing is given by:

(10) The requests are granted *according to the* first-come-first -served rule in any part of the system.

B. Topic Elaboration/clarification

The second subcategory of discourse organising bundles is used to elaborate and clarify topics in ERAs. The bundle *according to the* (136) in extract (10) communicate the method of granting requests. In the findings, the bundle *in order to (358)* is used as a marker to expand/ give more information to the reader (see extract 11). In extract (12), the bundle on the other hand (86) signals contrast and comparison.

(11) This paper proposes a new methodology that combines the mathematical models of protection system performance and of UBP *in order to* adequately solve the fault section estimation problem.

(12) The NoC architectures having complete buffer elimination used to be simple and power-efficient, but *on the other hand*, faced performance issues.

3.3.2.3 Stance bundles

Stance bundles are used to convey the writer's attitudes and ideas. They also refer to knowledge status of the information in the following proposition: certain, uncertain, or probable/possible' (Biber et al., 2004: 389). It is noticeable that there are only a few uses of stance bundles (6%) in the ERAs corpus, and the patterns of stance bundles are clausal. Interestingly, structural trends reflect the use of the functions of the bundle in this function type. As can be seen from extracts 13–15, non-human subject pronouns (e.g., it) are employed by the writers. In the same vein, writers use stance expressions to convey their ideas on objects.

Based on these findings, epistemic stance that focuses on impersonal expression can be seen, e.g., *to the fact that (24)**. From extract (13), it is evident that the bundle *to the fact that* (24) implies a degree of certainty regarding the information being discussed. Contrasting with the word *probably* that it follows in the sentence, which introduces ambiguity, the phrase *to the fact* implies certainty and presents a factual representation. It is important to note that this may result in a contradictory context: in the event that *the fact that* is accompanied by the word *probably*, a contradiction is created. The writer may be

certain about certain aspects (as indicated by *the fact that*) and acknowledge uncertainty in other respects (as indicated by *probably*).

> (13) This is probably attributed *to the fact that* for the less reactive MEA, more of MgO hydrates at late age, during which less expansive stress can be relaxed due to the viscous deformation of the cement matrix, leading to larger expansion.

> (14) *It should be noted* that the quantitative effects reported in this study are dependent on the individual cement, silica and PCE compositions.

> (15) *It can be* seen that the experimental results are in good agreement with the results obtained theoretically by using Eq. (7).

In extract (14), the bundle *it should be noted* (26) serves multiple functions in academic discourse. The phrase signals to the reader that what follows is noteworthy and requires attention. It prepares the audience for information that is likely to be crucial for understanding the topic at hand. By using this phrase, the writer underscores the significance of the information, emphasising its relevance within the broader context of the discussion. Furthermore, *it should be noted (26)* serves as a discourse organiser. It acts as a transitional device, helping to move from one idea to another while maintaining coherence in the text. The phrase bridges different sections or points of discussion.

While in extract (15), the use of the bundle *it can be (171)* conveys the ability to do the action and it includes an engagement feature that involves the reader directly.

The bundle *the heat transfer* (1%) is an example of a noun phrase (NP) that reflects specific content in the field of engineering. This bundle is particularly relevant to technical contexts and is exemplified in extract (16) below**.**

> (16) Kelly and Swenson **[**10**]** studied *the heat transfer* and pressure drop characteristics of a splash grid type of cooling tower packing**.**

3.3.2.4 Multi-functional bundles

As can be seen in the findings of the study, lexical bundles play a significant role in shaping academic writing through their multi-functional nature. The analysis highlights that bundles such as *it should be noted* (26) and *at the same* (81) serve multiple purposes depending on the context. These bundles function as stance markers to express certainty or highlight important points, discourse organizers to ensure logical connections between ideas, and tools for linking information clearly within the text. This multi-functionality emphasises their importance in enhancing both readability and coherence in academic discourse.

For example, the bundle *it should be noted* (26) is used in engineering research articles for multiple purposes. At the surface level, it signals the importance of what follows, helping the reader focus on key information. At a deeper level, it acts as a discourse organizer, connecting different ideas and ensuring the text flows smoothly. In extract (14), *it should be noted* (26) facilitates transitions between points, making it easier for the writer to explain their argument clearly.

The bundle *at the same* (81) also demonstrates its ability to serve different purposes. In extract (6), *at the same* (81) functions as a prepositional phrase to refer to location or quantity. It can also indicate time or serve as a reference. For instance, *at the same time* can both describe a moment and connect it to other parts of the sentence. This makes the writing clearer and helps the reader understand how the ideas are interrelated. In this regard, these bundles often indicate logical relationships and help link different parts of a text coherently. This is especially useful when explaining diagrams or reporting lab procedures, where it is crucial to show the flow of ideas and how they are connected. For instance, bundles like *it should be noted* (26) and *at the same time* make it easier to follow the writer's explanations and understand the steps being described.

It should also be noted that the multi-functional nature of lexical bundles reflects the need for awareness when writing or reading such materials, particularly in an engineering context. For example, the bundle *it should be noted* (26) serves as a good reminder for students to be cautious. This bundle frequently functions as a discourse

organizer, providing additional information or clarification to ensure the information is presented logically and clearly. By understanding the various functions that bundles such as *it should be noted* (26) can perform depending on the context, students can improve both their reading comprehension and writing skills in technical and academic settings.

3.4 Specific lexical bundles in the ERAs corpus

Previous research has not sufficiently addressed the characteristics of lexical bundles that are found in the engineering domain **(**cf**.** Cortes, 2004; Hyland, 2008**).** In this section I decided to provide list of the lexical bundles that are specific to engineering by looking at frequency and dispersion, these are the *Bottom 50***.** Pedagogically, these lexical bundles**'** inventory can be integrated with frequent lexical bundles **(**Top 50**)** when teaching engineering students**.** I do not include the lexical bundles that are in the middle range [20]of the lexical bundles' findings since their structure and function are similar to the Top 50 three-word and four-word bundles.

I set criteria for the Bottom 50 3-word and 4-word bundles**.** Section 3.4.1 discusses the characteristics of less frequent lexical bundles based on set frequency and range distribution, while Section 3.4.2 looks at the discourse functions that the Bottom 50 lexical bundles perform in ERAs.

3.4.1 Structure of specific lexical bundles in the ERAs corpus

As illustrated in Table 11, there are three main structures of the Bottom 50: incomplete noun-phrases (52%), incomplete dependent clauses (21%), verb phrase fragments (12%), and other expressions (15%). The three structures are based on previous studies of lexical bundles' structure from Biber et al. (2004). The frequency of bundles from the Bottom 50 ranged from 13 to 2[21], and they were all dispersed over 2 out of 100 engineering research articles. The criteria included getting rid of interrupted words or symbols (§ 3.2.3).

---

[20] The middle range three-word and four-word bundles have similar structure and function as the Top 50 three-word and four-word bundles, such as *the sum of the (10), this could be (9), in the production of (6), was reported that (6).*

[21] It should be noted that when I selected the Bottom 50 three-word and four-word bundles that have the frequency at 2, it automatically gave me the lexical bundles that have the frequency ranging from 2 to 13,

Structural characteristics					
Number of bundle types	**Incomplete noun phrase**	**Incomplete dependent clauses**	**Verb phrase fragments**	**Others**	**Total**
Number of 3-word bundle(Percentage)	25(50%)	10(20%)	6(12%)	9(18%)	50(100%)
Number of 4-word bundle (Percentage)	27(54%)	11(22%)	6(12%)	6(12%)	50(100%)

Table 11 Distribution of Bottom 50 3-word and 4-word bundle types by structure

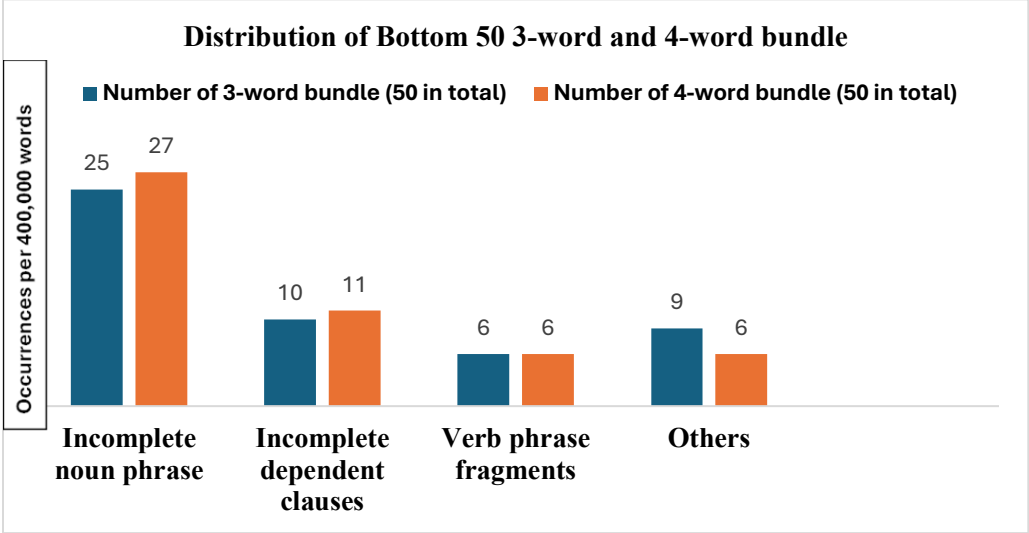


Figure 3 Overall distribution of Bottom 50 lexical bundle in ERAs: Structural type

Table 12 represents the main structures of the Bottom 50 3-word and 4-word lexical bundles. First, the incomplete noun-phrase bundles type (65%) comprises six subcategories. For example, NP based on technical terms in engineering (e.g.,*yield stress values (2)*), NP fragments with coordinated binomial phrases (e.g.,*first and the second (2)*), NP fragments with of-phrase fragments (e.g.,*efficiency of the expander (10)*), NP fragments with other prepositional fragments (e.g.,*zone between the (4)*), NP fragments with to-clause

and they all dispersed at 2. Since the frequency range at 2 of the Bottom 50 three-word and four-word bundles is fairly wide, there more than 10, 000 lexical bundles that have the frequency at 2.

fragments (e.g.,*yield stress to (2)*), and NP fragments with ing-clause fragments (e.g.,fluid leaving the turbine (13)).

Structural types of LBs in ERAs corpus	Numbers of LBs (Percentage) (N=100)	Example
1. Phrasal: Lexical bundles that incorporate incomplete noun phrase	52 (52%)	
1.1. Noun phrase fragments with coordinated bionomal phrase		*fly ash and (8), first and the second (2)*
1.2. Noun phrase fragment with of-phrase fragment		*influence of thermal (3), efficiency of the expander (10)*
1.3. Noun phrase with other prepositional fragment		*flow resistance in (2), flash evaporation from a (2)*
1.4. Noun phrase with to-clause fragment		*yield stress to (2), first to use the (2)*
1.5. Noun phrase with with ing-clause fragment		*fluid exiting the (3), fluid leaving the turbine (13)*
2. Lexical bundles that incorporate verb phrase fragments	12 (12%)	
2.1.Verb phrase (non-passive verb)		*yields the following (3), defined three types of (2)*
2.2.Verb phrase with other prepositional phrase fragments		*reflected in the (2), flow into the (2)*
2.3. Verb phrase with to-clause fragment		*flows to the condenser (2)*
3. Lexical bundles that incorporate incomplete dependent clauses	21(21%)	
3.1. Noun phrase fragments with passive verb phrase fragment		*efficiency was found to (2), findings can be summarised (2)*
3.2. Participle verbs with prepositional phrases fragment		*briefly explained in (2), briefly summarised below (2)*
3.3. Noun phrase fragments with copula be/verbs		*flow can be (2), flashing tank is between (2)*
3.4. Noun with verbs and that-clause		*finding is that the (3), figure shows that the (2)*
4. Others	15 (15%)	*yield stress values (12), flow reaction turbine rotor (2)*
		fluid as it (2), fluid as it leaves (2)

Table 12 Structural types of Bottom 50 3-word and 4-word lexical bundles in the ERAs corpus

Table 12 represents the main structural types of the Bottom 50 3-word and 4-word lexical bundles. The first and most frequent type is Incomplete noun phrase, accounting for 52% of all lexical bundles. These fragments are divided into five subcategories. The first subcategory includes noun phrase fragments with coordinated binomial phrases, such as fly ash and (8) and first and the second (2), which are often used for comparisons or enumerations. The second subcategory consists of noun phrase fragments with of-phrase fragments, such as influence of thermal (3) and efficiency of the expander (10), which describe relationships or dependencies. The third subcategory includes noun phrase fragments with other prepositional fragments, such as flow resistance in (2) and flash evaporation from a (2), which indicate spatial or positional meanings. The fourth subcategory comprises noun phrase fragments with to-clause fragments, such as yield stress to (2) and first to use the (2), which express direction or purpose. Finally, the fifth subcategory features noun phrase fragments with ing-clause fragments, such as fluid exiting the (3) and fluid leaving the turbine (13), which describe ongoing actions or processes.

The second most frequent type, incomplete dependent clauses, represents 21% of the lexical bundles and is grouped into four subcategories. The first subcategory includes noun phrase fragments with passive verb phrases, such as *efficiency was found to* (2) and *findings can be summarised* (2), which describe results or summaries. The second subcategory features participle verbs with prepositional phrases, such as *briefly explained in* (2) and *briefly summarised below* (2), often used for concise descriptions. The third includes noun phrase fragments with copula be/verbs, such as *flow can be* (2) and *flashing tank is between* (2), which describe conditions or states. The final subcategory comprises noun phrases with verbs and that-clauses, such as *finding is that the* (3) and *figure shows that the* (2), which clarify interpretations or key findings.

The third type, verb phrase fragments, accounts for 12% of the lexical bundles. These are divided into three subcategories. The first includes verb phrases with non-passive verbs, such as *yields the following* (3) and *defined three types of* (2), which indicate actions or results. The second subcategory consists of verb phrases with other prepositional

fragments, such as *reflected in the* (2) and *flow into the* (2), which describe directional or positional relationships. The third subcategory includes verb phrases with to-clause fragments, such as *flows to the condenser* (2), which describe movement or direction.

Lastly, the *Others* category accounts for 15% of the lexical bundles and includes subject-specific terms and bundles that do not fit into the first three groups. Examples include domain-specific terms like *yield stress values* (12), which reflect technical usage, and other structurally diverse bundles such as *fluid as it* (2) and *fluid as it leaves* (2), which are process-related phrases that lack clear categorisation in the third structural group.

In summary, Incomplete noun phrases dominate the Bottom 50 bundles, highlighting their significance in engineering academic writing. The other structural types, while less frequent, also contribute to the clarity and precision needed to explain complex technical concepts effectively.

3.4.2 Functions of particular lexical bundles in the ERAs corpus

In order to identify the function of the Bottom 50 3-word and 4-word bundles, and to identify the function that the Bottom 50 lexical bundles perform in ERAs, the analysis of their functions is based on Biber et al.'s (2004) functional taxonomies: referential, discourse organisers, stance, and other.

Discourse functions					
Number of bundle types	**Referential**	**Discourse Organisers**	**Stance**	**Others**	**Total**
Number of 3-word bundle	22 (44%)	8 (16%)	3(6%)	17 (34%)	50(100%)
Number of 4-word bundle	19 (38%)	19 (38%)	2(4%)	10(20%)	50(100%)

Table 13 Distribution of Bottom 50 3-word and 4-word bundle types by function

As shown in Table 13, the structural trends of the Bottom 50 lexical bundles are reflected in their most frequent discourse functions. These functions predominantly rely on NP-based (noun phrase-based), PP-based (prepositional phrase-based), and clause-based

structures as their main forms (see Table 1). Figure 3 below illustrates the distribution of the Bottom 50 3-word and 4-word bundles in the ERAs corpus across different discourse functions. This distribution reveals that referential functions dominate in both 3-word bundles (22, 44%) and 4-word bundles (19, 38%), followed by discourse-organising functions, which are more prominent in the 4-word bundles (19, 38%) compared to the 3-word bundles (8, 16%). Stance-related bundles are minimal, with only three instances (6%) in the 3-word bundles and two instances (4%) in the 4-word bundles. The *Others* category includes 17 bundles (34%) in the 3-word group and 10 bundles (20%) in the 4-word group, reflecting a range of additional discourse functions. This pattern highlights the dominance of referential and organising functions in the corpus while maintaining an objective and technical focus, with limited use of stance expressions.

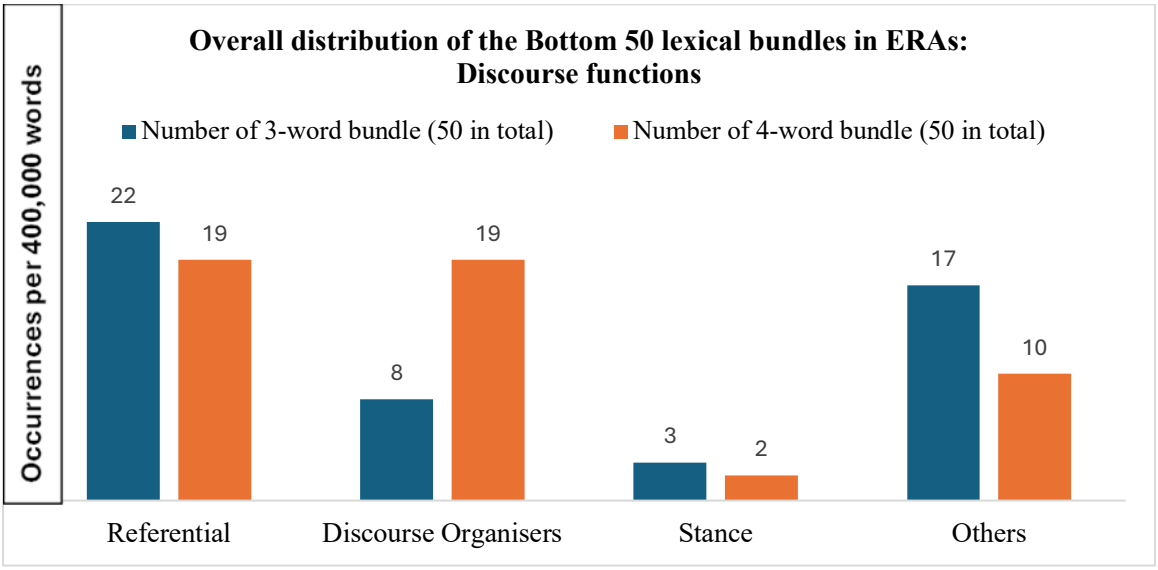


Figure 4 Overall distribution of the Bottom 50 lexical bundles in ERAs: Discourse functions

Category	Numbers of LBs (Percentage)	Example bundles from ERAs corpus
Sub-category	**(N=100)**	
1. Referential bundles		
A) Specific of Attributes	41(41%)	
A1) Quantity specification		*flux per unit (3), flow rate of air (2)*
A2) Tangible framing attributes		*efficiency of the expander (10), flow chart of (2)*
A3) Intangible framing attributes		*flow of air (2), defined in terms of (2)*
B) Time/Place/Text reference		
B1) Place reference		*flux in the (3), profile along the distribution (2)*
2. Discourse organising bundles		
A) Topic Introduction/Focus	27(27%)	*fluid leaving the turbine (13), first to use the (2)*
B) Topic Elaboration/comparison		*efficiency as compared to (3), flows through the (2)*
3. Stance bundles		
A) Attitudinal/Modality stance	5(5%)	
A1) Ability		*flow can be (2) flow rate can be (2)*
4. Others	27(27%)	*yield stress are (2), flow simple reaction turbines (2)*

Table 14 Functional types of the Bottom 50 3-word and 4-word lexical bundles in the ERAs Corpus

3.4.2.1 Referential bundles

Based on findings from the ERAs corpus, referential bundles account for 41% of all discourse functions and are grouped into two main sub-categories: specific attributes and time/place/text references**.** The ERAs corpus is a specialised corpus, and the bundles have specific attributes that refer to quantities or amounts in engineering discourse such as *flow rate of* (*2*), *flux per unit* (*3*).

- **Specific attributes**

Bundles in this sub-category are used to describe or quantify specific attributes, particularly in relation to uncountable nouns commonly encountered in engineering texts. For instance, the bundles *flow rate of* (*2*) and *flux per unit* (*3*) are frequently used to introduce and quantify physical parameters, as demonstrated in the extracts 17 & 18 below.

(17) In its most common use, both the pressure gradient and the *flow rate of* air vary during the test**.** This greatly simplified the measurement, but prevents the use of basic equations for the flow of air through a bed of packed particles**.**

(18) Where q is the heat *flux per unit* area, h is the convective heat transfer coefficient, s is the Stephane Boltzmann constant and T0 is the ambient temperature**.**

The bundles in this sub-category, such as *flow rate of* (2) and *flux per unit* (3), as shown in extracts 17 and 18, not only describe or quantify attributes related to uncountable nouns commonly found in engineering but also demonstrate their role in quantifying technical parameters. However, it is important to consider that the presence of these particular bundles may reflect the specific topics of the research articles included in the corpus. This relates to the issue of representativeness, as the ERAs corpus was designed to reflect selected areas of engineering research rather than the full range of engineering disciplines.

The bundles in this category describe the characteristics of technical nouns commonly used in engineering discourse. For example, the bundle *flow rate of air* (2), in extract (19), describes a measurable physical property, while *efficiency of the expander* (10), in extract (20), refers to an abstract characteristic related to system performance. Although such bundles may not be generalisable across all academic disciplines, they are

highly relevant in English for Specific Purposes (ESP) contexts. Specifically, they support learners in engineering by providing precise language for describing complex processes and technical components.

(19) In its most common use, both the pressure gradient and the *flow rate of air* vary during the test.

(20) The volumetric *efficiency of the expander* is highly dependent on the rotational speed.

- **Time/place/text reference**

Referential bundles in this sub-category are used to denote time, place, or references to specific elements within the text, such as figures, tables, or spatial positions. For instance, the bundle *zero at the* (2), as shown in extract 21, refers to a specific spatial location — *the origin point* — indicating where the temperature or measurement begins. This usage supports spatial referencing commonly found in technical descriptions.

(21) As shown in the Figure, the heat flux is initially *zero at the* origin point, and then increases dramatically as the film thickness increases**.**

(22) The main disadvantage is its generalization incapacity and the graphical representation of the protection system is *difficult to manage when* applied to large electric power systems.

The use of referential bundles in the ERAs corpus demonstrates their importance in both quantifying technical parameters and providing precise references to time, place, and textual elements. This dual role enhances clarity and coherence in the technical communication required in engineering research articles

3.4.2.2 Discourse organisers

Discourse organising bundles (27%) can be divided into two major functions: topic introduction/focus and topic elaboration/clarification.

- **Topic Introduction/focus**

The topic introduction bundle *first to use the* (*2*) in the ERAs corpus is used to introduce a new topic or focus in the discussion, as shown in extract (23):

(23) This study is the *first to use the* Van der Pol Oscillator for weak signal detection**.**

This bundle is typically employed to highlight novelty or originality in research, drawing the reader's attention to an innovative method, approach, or perspective being discussed.

- **Topic Elaboration/clarification**

The second function of discourse organising bundles is to provide elaboration or clarification on a topic. The bundle *briefly explained in* (2) functions as a discourse organiser within academic texts. In extract (24), it signals to readers that further details are provided in subsequent subsections, serving as a roadmap for navigating the content. This is particularly important in research writing, where clarity and structure help readers follow complex information. As such, discourse organisers like this bundle contribute to communicative effectiveness by guiding the reader through the text. Recognising their role supports the teaching of English for Specific Purposes (ESP), where learners need to develop skills for organising and presenting academic content clearly.

(24) Each part is *briefly explained in* different subsections**.**

Similarly, the bundle *efficiency as compared to* (*3*) in extract 25 is frequently used as a marker to clarify or expand upon the topic, particularly when contrasting results or methods. The use of this subcategory depends on the context and the writer's intention to clarify complex points.

(25) Their results indicated that the composite TE device shows a 24**.**8**%**, 26**.**2**%**, and 29**.**9**%** increase in conversion *efficiency as*

compared to a conventional TE device when the hot surface temperature T h ¼ 550 K, 450 K and 350 K respectively

Moreover, the bundle *efficiency as compared to* (*3*) is used to draw a comparison, elaborating on the significant improvements observed in the results. This helps the reader better understand the context and relevance of the findings. Such bundles are instrumental in providing detailed explanations or supporting key arguments within the text.

Therefore, discourse organising bundles in the ERAs corpus serve an essential role in structuring information. The topic introduction/focus function ensures a clear presentation of new ideas or innovations, while the topic elaboration/clarification function aids in deepening the reader's understanding of complex topics. Together, these bundles enhance the coherence and accessibility of technical writing.

3.4.2.3 Stance bundles

Stance bundles reflect the writer's attitude towards actions or ideas presented in the text. These bundles often convey certainty, possibility, or obligation, allowing the writer to position themselves within the discourse while guiding the reader's interpretation. In the ERAs corpus, stance bundles account for 5% of all bundles and typically involve modality to express levels of certainty or uncertainty. Stance bundles also serve an important directive function in technical writing. They are used to emphasise important points, guide readers through new arguments, and draw attention to key findings or facts. This helps focus the reader's cognitive processes on critical information, enhancing understanding and retention of technical material.

One example of the specific bundle found in the ERAs corpus is *flow rate can* (2), which is used to express technical possibilities or changes. This bundle features a non-human subject (flow rate), reflecting the objective and formal style typical of engineering texts. Moreover, in extract (26), the bundle *flow rate can* (2) demonstrates how modality (can) is used to highlight potential actions, providing certainty while maintaining a neutral stance:

(26) While the air flow rate is circulated counter-flow by an axial fan and is maintained at frequency switcher, the water mass *flow rate can* be changed manually by means of a balancing valve**.**

Here, the bundle focuses the reader's attention on the operational capabilities of the system while preserving objectivity. The use of *can* expresses certainty about what is technically feasible, guiding the reader's understanding of the described processes.

3.4.2.4 Other bundles in engineering

Other bundles in this category are those that do not well fit into the main functional groups. These bundles are primarily NP-based (noun phrase-based) and PP-based (prepositional phrase-based), reflecting their strong alignment with engineering-specific content. For example, the bundle *flow simple reaction turbines* (*2*) is NP-based and directly references engineering concepts, as demonstrated in extract (27):

(27) Radial outward *flow simple reaction turbines* exhibit a centrifugal pumping effect as discussed by a number of researchers **[**24, 25**].**

In addition, other bundles in this category often describe the movement of substances, such as fluids, gases, or devices, within engineering processes. A clear example is the bundle flow through the (2), which highlights the transmission or passage of energy or substances. This is illustrated in extract (28):

(28) Most of the transmitted energy *flows through the* shorting resistor**.**

These bundles are specifically used to convey dynamic processes or operations that are central to engineering discourse. Their emphasis on movement, transmission, and action reflects the technical nature of the corpus and the importance of describing mechanical or physical processes with precision.

The Other bundles category highlights the unique use of NP-based and PP-based bundles to describe engineering-specific content, particularly dynamic actions like movement or

transmission. These bundles play an important role in technical communication by ensuring clarity and specificity in describing engineering processes.

3.5 Overall Patterns of the Top 50 and Bottom 50 Lexical Bundles in Engineering

The findings reveal distinct yet complementary patterns in the structural and functional characteristics of the Top 50 and Bottom 50 lexical bundles in the ERAs corpus, reflecting the specialised nature of engineering discourse. Structurally, the most frequent bundles across both groups are prepositional phrase fragments (PP) (34 out of 100), noun phrase fragments (NP) (28 out of 100), which highlight the field's reliance on precise technical terms and relationships between concepts. Verb phrase-based (VP-based) bundles (26 out of 100) are the second most frequent type, often used to describe actions or processes, while clausal fragments (8 out of 100) and others (4 out of 100) appear less frequently, suggesting that engineering texts favour factual descriptions over complex argumentative structures. Functionally, referential bundles (50 out of 100) dominate the corpus, aligning with Biber et al.'s (2004: 384) observation that referential expressions are key to identifying and describing physical or abstract entities or their attributes. Discourse organisers (35 out of 100) are the second most frequent, serving to guide readers and structure information clearly, with examples like *first to use the* (*2*) introducing novelty or important findings. Stance bundles (14 out of 100) are less frequent but express certainty or possibility, as seen in *flow rate can* (*2*), while other (1 out of 100) serve specialised roles. These patterns resonate with prior studies by Hyland (2008) and Rezoug & Vincent (2018), which highlight the structural and functional roles of lexical bundles in engineering sub-disciplines, particularly those related to quantification and process description. The study also underscores the importance of incomplete clause bundles in the Bottom 50, reflecting the fragmentary nature of technical language in engineering. Compared to disciplines like the humanities, where stance and referential bundles are more frequent to convey evaluations and relationships, engineering prioritises non-human subjects and text-organising bundles to ensure clarity and precision. These findings not only align engineering discourse with broader academic writing conventions but also emphasise the

need for engineering learners to develop an awareness of these unique linguistic patterns to enhance their reading and writing in professional and academic contexts.

In addition to these overall patterns, further analysis reveals how the function of lexical bundles relates to their structural form and length. The frequency patterns of lexical bundles reveal a strong relationship between their functions and structural characteristics. Noun phrase and prepositional phrase fragments are the most frequent, constituting 32% of three-word bundles and 30% of four-word bundles, reflecting their predominant role in referential functions, where they provide specific information and descriptions. Verb phrase fragments account for 12% of three-word bundles and 14% of four-word bundles, playing a significant role in stance functions, allowing authors to express evaluations or judgements. Clausal fragments, though less frequent (3% in three-word bundles and 5% in four-word bundles), primarily serve as discourse organisers, connecting larger textual units. Notably, three-word bundles exhibit a balanced distribution between referential (25%) and discourse-organising functions (35%), whereas four-word bundles show a stronger preference for referential functions (25%) and a decline in discourse-organising functions (14%), suggesting that longer bundles are more suited for conveying complex referential meaning rather than managing discourse flow. These patterns align with findings in prior research (e.g., Biber et al., 1999), lexical bundles show flexibility in both their structure and function, allowing them to serve various purposes in expressing meaning and organising discourse.

3.6 Summary

This chapter has presented a summary of analytical and theoretical frameworks for lexical bundles as the methodology of this study. The tools and equipment for analysis have been discussed, starting with how data collection was set up. I have detailed the reasons why I choose ERAs as the main data set (see also Chapter 1) for pedagogical reasons, as well as the processes involved in data selection, the coding process and in particular the criteria for lexical bundles' structural and functional categorisation. The categories used for analysis are those developed within the lexical bundles taxonomies of Biber et al. (2004)

for lexical bundles' structure and function in academic registers. Subsequent chapters will address the initial research questions stated in Chapter 1. The development of materials based on lexical bundles' structure and functions will be addressed through the process of developing a workshop for teaching lexical bundles to engineering students in the Chapter 4.

4. Chapter Four: Teaching lexical bundles to engineering students: methods, results and discussion

4.1. Introduction

This chapter draws on the structural and functional categorisation of lexical bundles that was introduced in Chapter 3. This section focusses on the application of lexical bundles in the instruction of engineering students by selecting the Top 50 most frequently used 3-word and 4-word bundles in ERAs. These bundles include *in order to* (358, 74), *due to the* (276, 78), *the number of* (202, 54), *the use of* (161, 59), and *the effect of* (139, 42) (refer to Section 4.2.1). The objective of this chapter is to address the research questions concerning the integration of the data-driven learning (DDL) approach into instruction and the use of lexical bundles in teaching.

The chapter discusses methods for teaching lexical bundles, such as designing activities and using concordances (§ 4.2). It also describes the implementation of teaching lexical bundles in a workshop (§ 4.2.4.2). Furthermore, this chapter discusses data collection during the teaching of lexical bundles, including the use of research instruments and the evaluation of inter-rater reliability (§ 4.2.5). It investigates whether the workshop helped students improve their knowledge of lexical bundles to support their ability to write engineering research articles (§ 4.3). Specifically, it evaluates students' understanding of lexical bundles through the activities provided (§ 4.3.1) and analyses grammatical errors related and unrelated to lexical bundles in their pre-test and post-test writing (§ 4.3.2.1 & 4.3.2.2). This analysis highlights the strengths and weaknesses of using this teaching approach with the students (§ 4.3.3 & 4.3.4).

To address the research questions, this chapter examines how to teach lexical bundles, how to use the ERAs corpus in instruction, and how this method supports engineering students in gaining knowledge about lexical bundles. It also discusses the challenges and limitations of teaching these bundles effectively.

This study involves three key components. First, the ERAs corpus was compiled and analysed to identify the most frequent and functionally relevant lexical bundles used in engineering research articles. Second, these targeted lexical bundles were selected and

integrated into instructional materials for teaching, allowing students to explore their structure, meaning, and use in academic writing. Third, data were collected during the teaching process through classroom observations, student tasks, and feedback to evaluate how well students understood and applied the bundles. Each of these stages is explained in detail in the following section.

At the outset of this chapter, it is important to note that the teaching task focused on five of the top 50 lexical bundles identified in the corpus. This decision was made because these bundles are highly frequent and relatively straightforward to teach, making them more accessible for students with varying levels of proficiency. While less frequent bundles were discussed in Chapter 3, their more complex structures posed significant challenges for effective teaching in the short time frame available. Future research might explore incorporating less frequent bundles into teaching activities to examine their potential impact, particularly as these bundles often reflect discipline-specific language features. However, for this study, the focus was intentionally placed on bundles that are most common and accessible to learners.

4.2. Methodology

Three aspects are involved in designing the methods of the study, and these are summarised in Table 15 below.

Aspect	Methods	The selection of methods
1. Selection of target lexical bundles **2. Teaching of target lexical bundles**	• Frequency, dispersion range • Data-driven learning (DDL) approach • Principles of materials development	Biber et al., 1999; Biber et al., 2004 Johns, 1991; Tomlinson, 2011; Cortes, 2006; Lewis, 2000;

3. Data collection	• Teaching of lexical bundles • Types of exercises • Introduction of AntConc concordance in the classroom • Pre-test and Post-test • Student interviews • Classroom recordings • Field notes • Students' notes • Inter-rater reliability	McEnery and Wilson, 2001; Hunston, 2002; Sripicharn, 2004; Sripicharn, 2010; Charles, 2015; Anthony, 2016 Mackey and Gass, 2005; Cortes, 2006; Creswell, 2014

Table 15 Methods used in teaching lexical bundles

First, the methods used to selecting target lexical bundles based on frequency and dispersion range were devised by drawing on Biber et al.'s (1999, 2004) lexical bundles' structural and functional taxonomies. The target lexical bundles are identified and classified according to their functions (§ 3.3 in Chapter 3).

Second, the teaching of lexical bundles was planned, including approaching the teaching of lexical bundles found in the ERAs corpus, designing materials and teaching lexical bundles in the classroom. Following Johns' (1991) DDL approach, students are also researchers in the ways that they explore data on their own. However, the degree of DDL use in the classroom depends on the context of this research, i.e., students' background. This study takes into account the students' background knowledge of corpora and lexical bundles. Thus, this study does not rely on the teacher fully initiating students into activities or maintaining complete control over the learning process. Instead, it encourages a balanced approach where students actively engage in tasks with some level of autonomy,

guided but not dominated by the teacher. Instead, the teacher acted as a facilitator to monitor or offer help whenever the students needed it *(*§ 4.2.3.5*).*

In terms of preparing and teaching students how to use concordances in the classroom, I have followed the principles described by Hunston *(*2002*)*, Sripicharn *(*2004, 2010*)*, Charles *(*2015*)* and Anthony *(*2017*)*. This study introduced the use of AntConc to the students in order for them to work with concordances and target lexical bundles.

This study embeds the principles of developing materials by drawing on Tomlinson (2011), who emphasises encouraging thinking and providing interesting activities. Tomlinson's (2011) principles of material development emphasise learner-centered, engaging, and authentic activities that encourage critical thinking. These principles align with Data-Driven Learning (DDL) methods as proposed by Johns (1991), which focus on empowering learners to explore and analyse language patterns independently using real-world data. When applied to the teaching of lexical bundles in Engineering Research Articles (ERAs), these principles and methods foster an environment where learners actively engage in critical thinking through hands-on activities.

A key activity implemented in the session involves students using AntConc, a corpus analysis tool, to explore and discover the structure and function of lexical bundles. This activity exemplifies the learner-centered approach advocated by Tomlinson. By interacting directly with corpus data from authentic ERAs tailored for engineering students, learners are tasked with identifying patterns and understanding the contexts in which lexical bundles are used. This process requires students to hypothesize, analyse, and draw conclusions, which are core components of critical thinking. Unlike traditional teaching methods where language rules are presented explicitly, this activity challenges students to uncover patterns themselves, fostering deeper engagement and analytical skills.

The activity also requires students to interpret the meaning of lexical bundles in relation to their specific contexts. For example, students must determine how lexical bundles such as *in order to*, *due to the*, or *the number of* function within authentic texts and decide on meanings that best align with their use in engineering discourse. This task involves evaluating multiple possibilities and considering both linguistic and disciplinary

contexts, which promotes higher-order thinking skills such as evaluation, synthesis, and application. By actively making these decisions, students practice critical thinking in a real-world linguistic context, preparing them for tasks they may encounter in academic or professional settings.

Moreover, the use of authentic ERAs ensures that students engage with materials that are directly relevant to their field of study. Authentic materials, as highlighted by Tomlinson, provide learners with opportunities to encounter language as it is naturally used. This authenticity adds an additional layer of complexity to the task, as students must navigate real-world language data that may include variations and nuances not typically found in simplified materials. This exposure encourages critical engagement with language, requiring students to analyse and adapt their understanding based on the evidence presented in the corpus.

The integration of technology, such as AntConc, plays a crucial role in extending learning beyond the traditional classroom. By using a corpus analysis tool, students develop technical skills alongside linguistic knowledge, enabling them to independently explore and analyse large datasets. This self-directed learning process mirrors inquiry-based approaches, where students take ownership of their learning journey. The ability to independently explore and interpret data aligns with Tomlinson's principles of encouraging thinking and creating meaningful activities. It also reflects the critical thinking process, as students must navigate unfamiliar tools, analyse complex data, and apply their findings to practical tasks.

In summary, the activities embedded in this session effectively operationalise Tomlinson's principles to promote critical thinking. Through the use of AntConc, students engage in hands-on exploration and discovery, analysing authentic ERAs tailored to their field. Tasks that require interpreting the meaning and function of lexical bundles encourage analytical and evaluative thinking, while the integration of technology fosters independence and inquiry-based learning.

These activities, grounded in both Tomlinson's and DDL's principles, create a rich learning environment where students actively develop critical thinking skills essential for

academic and professional success. This study integrates the principles of material development proposed by Tomlinson (2011), who highlights the importance of fostering critical thinking and engaging students through stimulating activities. The approach to material development outlined by Tomlinson (2011) aligns closely with the data-driven learning (DDL) methodology introduced by Johns (1991), as both emphasise creating activities that encourage active thinking and learner autonomy. It should be noted that this study does not focus on tasks in the pre-test and post-test. Therefore, the calculation of task confidence is not relevant to this study.

Task confidence in language testing refers to the belief that learners have in their ability to successfully complete language-related tasks. As defined by Koteková (2013), task confidence involves assessing how well language understanding models and learners can perform under various conditions, and how confident they are in their abilities. This concept highlights the interaction between learners' perceived capabilities and their actual performance in language-related activities.

Instead, I exploited the ERAs corpus in designing activities that are based on using lexical bundles in the ERAs corpus, and using the principles of activity design in the lexical approach from Hill et al. (2000). The pre-test and post-test are based on finding the correct lexical bundles included in the ERAs corpus (cf. Hill et al., 2000). The teaching procedures and materials used in this study were designed to support both linguistic competence and critical thinking skills among Thai engineering students. For instance, one of the materials used was a worksheet focusing on lexical bundles in engineering contexts. This worksheet included exercises where students analysed lexical bundles in authentic research articles, identified their functions, and discussed their implications for professional communication. By encouraging students to reflect on the use of these lexical bundles in various contexts, the materials aimed to enhance both their understanding and application of the terms. Furthermore, teaching procedures included group activities where students collaboratively created their own lexical bundles for engineering-specific scenarios, followed by peer evaluation. These activities fostered critical thinking by requiring students to assess the appropriateness and clarity of lexical choices in specific engineering contexts.

Additionally, discussions during class sessions were structured to prompt analytical thinking, such as questioning why particular lexical bundles are commonly used in specific sections of research articles (e.g., introductions vs. conclusions) and exploring their effectiveness in conveying information.

Third, data collection includes both the research instruments and the methods used for analszing the collected data (cf. Creswell, 2014). To ensure reliability, the inter-rater reliability of students' writing in the pre-test and post-test was assessed by applying the guidelines provided by Mackey and Gass (2005).

Considering the methods shown in **Error! Reference source not found.** above, they are used to guide me through the process of teaching lexical bundles to engineering students. I have divided implementing the teaching of lexical bundles to engineering students into four main stages: selection of target lexical bundles (4.2.1), materials development (4.2.2), implementing the workshop (4.2.3), analysis of students' use of lexical bundles (4.2.4). The research instruments and inter-rater-reliability employed in this study are explained in Sections 4.2.4.1 and 4.2.4.2. Each of these stages is described in detail in the following sections.

4.2.1. Target lexical bundles

Five bundles from the top 50 3-word bundles were used in the instruction. They were selected because they are frequent and are dispersed across 100 engineering articles. The five bundles, with their frequency and dispersion range, are: *in order to* (358, 74), *due to the* (276, 78), *the number of* (202, 54), *the use of* (161, 59), *the effect of* (139, 42) (see Table 4.2 below). In terms of familiarity, these five bundles selected from the top 50 3-word bundles were considered to be general and appropriate to the engineering students' background knowledge and level of English proficiency in order to understand the bundles used in the workshop. The five lexical bundles chosen for teaching in this study were selected primarily for their high frequency and broad applicability in academic writing. These bundles are common not only in engineering research articles but also across other academic disciplines, which supports their utility in developing general academic writing skills. However, it is recognised that this approach may limit the focus on discipline-

specific language features. An alternative approach could involve selecting bundles based on their specific functions or structural types, or even targeting those that students frequently misuse or underuse. Future research could explore these approaches to provide deeper insights into how teaching less frequent or more discipline-specific bundles might impact students' writing in engineering contexts. For this study, the emphasis was placed on accessibility and immediate applicability to ensure effective teaching within the limited instructional time available.

Teaching the Top 50 three-word bundles provides students with foundational knowledge of lexical bundles. This foundation can be expanded through targeted activities that focus on analysing and applying specific lexical bundles from their field of study. For example, students can work with field-specific texts to identify and categorise bundles, understand their functions in context, and use them in their own writing. Collaborative tasks, such as peer reviews or group discussions, can also help students refine their understanding and application of these bundles. This gradual approach ensures that students effectively transition from general to field-specific lexical bundle knowledge.

Appendix 3 consists of 200 three-word and four-word lexical bundles, both from the top and bottom, which occur at least twice in two different texts within the ERAs corpus. These bundles have been normalised to a frequency per 400,000 words. Many of the bundles exhibit complex grammatical structures, including prepositional phrases *(e.g., in the case of),* noun phrases *(e.g., the results of the*), and verb phrases *(e.g., is shown to be)***.**

These structures often involve abstract or technical language, which can be difficult for students to understand and use correctly without significant practice. For instance, students may struggle with the syntactic flexibility of some bundles, such as those that require accurate use of prepositions or article-noun combinations. Additionally, the bundles frequently appear in highly specialised contexts, making it harder for students to grasp their usage without exposure to domain-specific texts. The analysis in Appendix 3 supports these observations, showing that students often misinterpret or underutilize these

bundles in their writing, indicating the need for additional scaffolding and practice. By addressing these challenges through targeted instruction, students can gradually build their competence in using these structures effectively.

Teaching the top 50 three-word bundles provides students with foundational knowledge of commonly used lexical bundles, which can then serve as a stepping stone for understanding more specific bundles used in their field. This is achieved by first familiarising students with these high-frequency bundles through activities such as contextual analysis, where students identify their functions in authentic texts. For example, students may learn that the bundle *it is important to* signals emphasis and is commonly used in academic writing. Once this foundational knowledge is established, students can analyse engineering-specific texts to identify additional bundles that are more specialised, such as *as shown in Figure 1* or *the results indicate that*, which are prevalent in engineering discourse. Further, students can engage in scaffolded learning activities, such as comparing and contrasting the use of general academic bundles with field-specific ones, to deepen their understanding. By gradually transitioning from general to specific bundles, students build their ability to recognise and use lexical bundles that align with their academic and professional needs. Additionally, exposure to field-specific bundles is reinforced through targeted reading exercises and collaborative writing tasks, where students apply their knowledge to create texts that mimic authentic engineering communication.

Moreover, the findings from the students' writing in the pre-test and post-test reveal that although the general lexical bundles selected are familiar to the students, they are not able to use all of the target lexical bundles in their writing correctly (§ 4.3.2). This justification is also based on the fact using lexical bundles correctly is a sign of proficient users and it takes time for students to learn this. Thus, the use of lexical bundles is considered problematic because they are not naturally acquired through regular language use. Lexical bundles, as fixed or semi-fixed multi-word expressions, require explicit teaching and practice to be fully understood and applied, particularly in academic or professional contexts. This issue is especially significant in the context of English for Academic Purposes (EAP) and English for Specific Purposes (ESP), where students need

targeted instruction to identify, understand, and use these bundles effectively. Without proper guidance, learners may struggle to grasp their structure and function, hindering their ability to communicate fluently in their respective fields (cf. Leńko-Szymańska, 2003; Cortes, 2004). This difficulty stems from the fact that lexical bundles are often context-specific and consist of sequences of words that are not easily identifiable as single, discrete units of meaning. Unlike individual words or common collocations, lexical bundles are typically learned through repeated exposure in specific academic or professional contexts, making them challenging for learners who lack consistent access to such environments. Biber et al. (2004) highlight that lexical bundles are not inherently intuitive for learners because they do not always conform to standard grammatical rules or convey clear standalone meanings. For instance, bundles like *on the other hand* or *in the case of* are functionally significant but require an understanding of their discourse and pragmatic functions. Furthermore, observations from this study revealed that students often struggled to recognise or reproduce lexical bundles accurately, even after initial instruction, indicating that explicit teaching and practice are essential. This suggests that without deliberate, targeted instruction and extensive exposure to authentic texts, students are unlikely to acquire these bundles naturally or use them effectively in their writing. In fact, learning lexical bundles has been found to be problematic for both native and non-native speakers, and it takes a long time to monitor and develop their learning (cf. Cortes, 2004). Therefore, only five lexical bundles from the Top 50 are selected in this stage.

4.2.2. Materials development

The purpose of this section is to produce materials according to the characteristics of lexical bundles, focusing on materials (§ 4.2.2.1), and piloting them before using them in the workshop (§ 4.2.2.2).

This section first defines the concepts of *materials* and *activities* as used in this study. *Materials* refer to instructional resources specifically developed or selected to support the learning of lexical bundles in engineering contexts. Examples include worksheets, handouts containing authentic lexical bundles, and excerpts from engineering research articles. These materials aim to provide learners with meaningful exposure to

lexical bundles and opportunities to practice using them effectively. *Activities* refer to classroom tasks and exercises designed to actively engage students in the learning process. These include analysing lexical bundles in authentic texts, group discussions on their functions, collaborative creation of lexical bundles tailored to specific engineering scenarios, and peer evaluations of these creations. The primary goal of these activities is not only to enhance linguistic accuracy but also to foster critical thinking skills, as students are required to apply and reflect on their knowledge in practical, discipline-specific contexts.

The materials for this study were designed based on the characteristics of lexical bundles, with a focus on their structure and functions. The primary objective was to help students understand lexical bundle structures and their practical applications. Using *AntConc* software, students gained hands-on experience both in class and outside the classroom. The aim was to enhance their learning experience by integrating *AntConc* as a tool for exploring lexical bundles. The content of the materials was tailored specifically for engineering students and included research articles sourced from the ScienceDirect database. These articles served as authentic materials, as students were already familiar with them through their use in gathering information for small engineering projects. By working with these authentic texts, students could relate the lexical bundles they studied to real-world engineering contexts.

4.2.2.1. Focus on materials: lexical bundles' forms and functions, AntConc concordances

The materials used in this workshop formed the foundation of the main activities and were designed to enhance students' understanding of corpus tools and lexical bundles in engineering research articles (ERAs).

The first session focused on introducing students to corpus and concordance concepts, as well as using the AntConc software to examine data within a small corpus of ERAs.

The second session built on this foundation by introducing the concepts of lexical bundles and engaging students in activities aimed at strengthening their understanding of the structure and function of lexical bundles in ERAs. This section provides a detailed discussion of the materials used in the classroom, including each activity and examples of student responses or outputs that demonstrate their understanding of lexical bundles. In particular, AntConc, a software tool for corpus analysis, can significantly enhance learners' engagement with lessons, particularly in language learning contexts.

The workshop was divided into two main sessions. AntConc is a software tool for corpus analysis, can significantly enhance learners' engagement with lessons, particularly in language learning contexts. The workshop was divided into two main sessions. can help learners engage with the lessons by enabling hands-on exploration of lexical bundles within authentic texts. This directly relates to the preceding discussion about the integration of authentic materials and activities. By using *AntConc*, students can independently analyse the frequency, structure, and function of lexical bundles, which complements their classroom discussions and collaborative activities. *AntConc* allows students to explore language patterns, deepening their understanding of language structures and usage. Additionally, by engaging with real-world text data, students can develop critical thinking and analytical skills, which are essential for language proficiency. This approach encourages active learning and critical thinking, as learners can see real-world applications of lexical bundles in engineering contexts and reflect on their practical usage. By presenting these materials and responses systematically, this section highlights how the activities contributed to the students' learning process.

Part 1 Introduction to ERAs Corpus

This subsection introduces students to the Engineering Research Articles (ERAs) corpus. It provides step-by-step instructions on how to upload and navigate the corpus using AntConc, enabling students to explore its content effectively. The primary focus is to familiarise students with the dataset and demonstrate its relevance to engineering writing.

The activities in this section are discussed in detail, accompanied by explanations of the students' learning processes and engagement.

Part 1**:** Getting to know ERAs corpus and concordance lines

Have you ever heard about corpus? What is a corpus?

A corpus (plural: corpora) is a collection of electronic texts. Corpora are built with a specific purpose in mind and are designed according to specific criteria. For example, if you want to check the language used in your field, you can build a corpus of relevant research articles written by experts. Corpora are accessed by using text analysis software. (Hunston, 2002; Charles, 2009). In the workshop, we will use a corpus named 'Engineering Research Articles Corpus' (ERAs corpus), which is compiled to help learn the use of lexical bundles from the engineering research articles.

Activity 1:

Focus: concordance lines

Purpose: to get to know the concordance lines

Instructions: Have you ever heard about corpus? Have you ever used concordance lines before? If not, scan the following lines 1 to lines 10 of the text taken from British Academic Written English Corpus **(**BAWE**).** Focus on the word 'house', and try to answer the questions that follow. You can discuss with the person near you. The answer will be explained and discussed together as a whole class.

Table 1*:* concordance lines 1 -10 taken from BAWE

N concordance			
1	what was 'really going on' at the halfway	house	- what was 'really going on' was residents
2	'one-stop shop' for advice on benefits,	housing	, education, lifestyle, etc. They also provide
3	nationality, education, economics, distribution,	housing	, labour protection and social security
4	classic' welfare state: health, education,	housing	, and the national insurance and national
5	responsibilities, which include: education,	housing	, planning applications, strategic planning
6	level 24. This includes improving education,	housing	, material circumstances, targeting young
7	would be used for investments in education,	housing	, health care, or job training. Contrary
8	inequalities in many areas, including education,	housing	, and employment. Ethnic minorities are
9	that I expect to look at how the feeds,	housing	, fertiliser, etc. can be improved to meet
10	and environmental factors such as income,	housing	, education, safety, access to fresh food

Questions:

1. What is the meaning of the word 'housing' in this context?

2.What types of words are commonly used in this context?

3. What type of word is most frequently used after ‘housing’? What are the other types of words used after ‘housing’?

4. What are the patterns of ‘housing’ used in the text?

5. In general, what is the meaning of the word house in the above concordances?

The possible answer of activity 1 is on the next page.

Possible answer:

1. The word ‘house’ or we call it as the ‘node’ word or the ‘target’ word.

2. Nouns

3. There are individual words or groups of words.

4. The house is frequently used as object in the sentences. The possible pattern (Line 6) for example, subject (This), Verbs (includes), Object (improving education, housing…)

5. The word ‘house’ means the normal house (in line 1), as you can see from the words that associated with the house e.g.,halfway, residents. Other meaning of the house referred to one of the government benefits systems (can see from line 2 to line 10). It can be noticed that the target word ‘housing’ used in the written words here seemed to be formal.

Activity 2:

Instructions: Notice the use of the word ‘house’ taken from British Academic Written English Corpus (BAWE) from Activity 1. Do you think the use of the word ‘house’ from Activity 1 differs from the use of the word ‘house’ in British Academic Spoken English Corpus (BASE), from Activity 2 ?

Table 2*:* concordance lines 1-10 taken from BASE

N concordance			
1	nineteenth century a novel called The Counting	House	[[voiced pause]] and by the way i use the
2	[[voiced pause]] in the [[voiced pause]] great	house	and then a younger servant a younger girl
3	it was within the compound of the great	house	surrounded by a high wall of Suffolk brick
4	Gladstone graveyard Anglican church plantation	house	my grandpa work so hard that he learn Latin
5	wood from the fence surrounding the great	house	or tools from the warehouse with the money
6	chapter from [[voiced pause]] this Counting	House	book right by the way i should say as a
7	[[voiced pause]] displaying them at in his	house	in London and allowing paying visitors
8	series [[voiced pause]] of discussions in the	House	of Commons through eighteen-fifteen and
9	estates with their rich and well equipped	houses	in the country and which was the worst
10	things like the woman being the angel in the	house	well most women in the nineteenth century

Now, you should know what concordance line is. Can you write the answer in your own words? Or you can write the answer in Thai. Then, we can share our answer on the board together. Then, check the meaning of concordance on the next page.

Possible answer of Activity 2: The target word ‘house’ was used in the BASE (from lines 1-10) means a normal house. You can see from the words that associated with the house e.g.,well-equipped, angel in the, his. As shown from lines 1-10, the word ‘house’ was normally used in the spoken language and it was more informal than the word ‘housing’ from Activity 1 above.

Possible Answer:

Concordance is a list of lines of text containing a node word, nowadays generated by computer as the principal output of a search of a corpus showing the word in its contexts and thus representing a sum of its usage. Also, we usually study a few words to the left and right position of the search term (Teubert and Cermakova, 2007, p. 104; Baker, 2007: p. 71).

You will learn how to sort the target words in the next section (see Part 2, Screen Shot 3: Concordance on the word ‘current’).

In Session 1, there were two main parts. The first part was to introduce the idea of corpora and concordance lines, while the second part involved using AntConc to learn about lexical bundles. From **Error! Reference source not found.**, the first part of Session 2 covered the concepts of corpora and concordances. The purpose was to prepare the students to use the ERAs corpus and AntConc concordances in studying lexical bundles’ forms and functions in Session 2. The concordance format used in the teaching task aimed to familiarize students with authentic examples of lexical bundles in context. However, it is acknowledged that the presentation of the concordance in this instance could have been improved for readability and teaching effectiveness. The gaps on either side of the node word, the variation in gap lengths on the left side, and the inclusion of related forms like "housing" instead of focusing solely on "house" may have made it more challenging for students to identify patterns. Additionally, the concordance was not sorted, and multiple examples were presented for some cases, which may have overwhelmed learners. For a first concordance, providing a single sentence for each case and sorting the examples could enhance clarity and usability. These considerations will be incorporated into future teaching designs to better support students’ pattern recognition and learning.

The phrasing of certain questions in the teaching task, such as *types of words* and *patterns of housing*, may have been too vague for students to understand clearly. To ensure better comprehension, these questions could be rephrased. For example, *types of words*

could be clarified as *different grammatical forms or word classes (e.g., noun, verb, adjective)* to provide clearer guidance for students. Similarly, *patterns of housing* could be revised to common uses of the word *housing* in different contexts *(e.g., housing as a physical structure versus housing as a social issue)*. These changes aim to make the task more explicit and meaningful. Additionally, during the teaching session, students were encouraged to ask questions if they needed further clarification about the activity instructions, ensuring their understanding of the task.

S2's reflections on Activity 1, titled getting to know ERAs corpus and concordance lines, offer valuable insights into her learning process and demonstrate the impact of this activity on her developing competence.

In aspect of effort required**,** S2 noted that understanding the contextual meaning of *house* was not straightforward. She highlighted the need to analyse surrounding words in each concordance line to determine the precise meaning and usage of the word. For example, she had to notice the words to the left and right in order to make assumptions or guesses about the meaning, function, or usage of the target bundles in a text. This process required effort and careful attention to detail, especially in recognising the role of collocates in shaping meaning.

In terms of learning outcome**,** S2 concluded that the activity successfully guided her understanding of *house* in different contexts. She found that focusing on surrounding words provided valuable insights into the meaning and function of *house* across various usages. This guided her preparation for subsequent tasks and activities.

In terms of developing competence, S2's ability to reflect on the role of surrounding words indicates her developing competence in using concordance tools to analyse lexical patterns systematically. She appreciated the scaffolding provided by the activity, which supported her learning process and helped her build foundational skills for more advanced tasks.

However, S2 experienced some initial difficulty in interpreting concordance lines, which highlights the cognitive demands of such tasks for students who are new to corpus tools. These challenges underscore the importance of providing adequate support and

guidance. As a teacher, there are several strategies that could enhance competence in such activities. Providing additional guidance through more explicit instructions or modelling examples could help students like S2 overcome initial difficulties. For instance, demonstrating how to analyse collocates step by step could provide clarity. Moreover, introducing simpler examples before moving to complex concordance lines might ease the learning curve, improve student confidence, and reduce cognitive load.

In the context of this study, learning refers to students' ability to understand, recognise, and effectively use lexical bundles in academic and professional writing. Learning can be measured through several indicators. First, students demonstrate understanding by identifying the structure and function of lexical bundles within authentic texts. For example, recognising that *in order to* introduces a purpose or goal reflects an understanding of its pragmatic function. Second, learning is evident when students can apply their knowledge of lexical bundles to produce original texts that are syntactically and contextually appropriate. This includes incorporating bundles such as *the results indicate that* accurately in their writing to convey findings or conclusions. Third, learning involves the ability to adapt and use lexical bundles in varied contexts, showing flexibility and a deeper comprehension of their usage. For instance, using a bundle like *on the other hand* to compare ideas in different argumentative contexts demonstrates advanced understanding. Additionally, qualitative feedback from peer evaluations and instructor assessments can provide further evidence of learning by highlighting improvements in students' awareness and application of lexical bundles over time. These indicators collectively define *learning* as not only the acquisition of knowledge but also the ability to apply that knowledge effectively in real-world academic and professional scenarios.

Finally, encouraging follow-up discussions where students share their interpretations in group settings could further reinforce their understanding and offer diverse perspectives on lexical usage.

To sum up, S2's reflections demonstrate that this activity effectively familiarised her with the use of concordance lines and enabled her to interpret the word *house* in context. The activity successfully built her foundational skills for more advanced tasks, despite

initial challenges. By refining the activity with enhanced scaffolding, gradual progression, and collaborative opportunities, it can better address student needs and further improve competence and confidence.

In Part 2, the AntConc tools were introduced, including how to use AntConc concordances to study lexical bundles (see Figure 9). Each step of loading and installing the program was clearly signposted to ensure ease of access. These activities were designed to prepare the students for working with AntConc. The students then followed step-by-step instructions to load, install, and explore AntConc concordances. This hands-on practice was intended to increase learner engagement and ensure that they could apply the tools confidently during subsequent lessons.

Part 2**:** Introduction to AntConc

Focus: Using AntConc to explore a small corpus of engineering research articles *(*ERAs corpus)

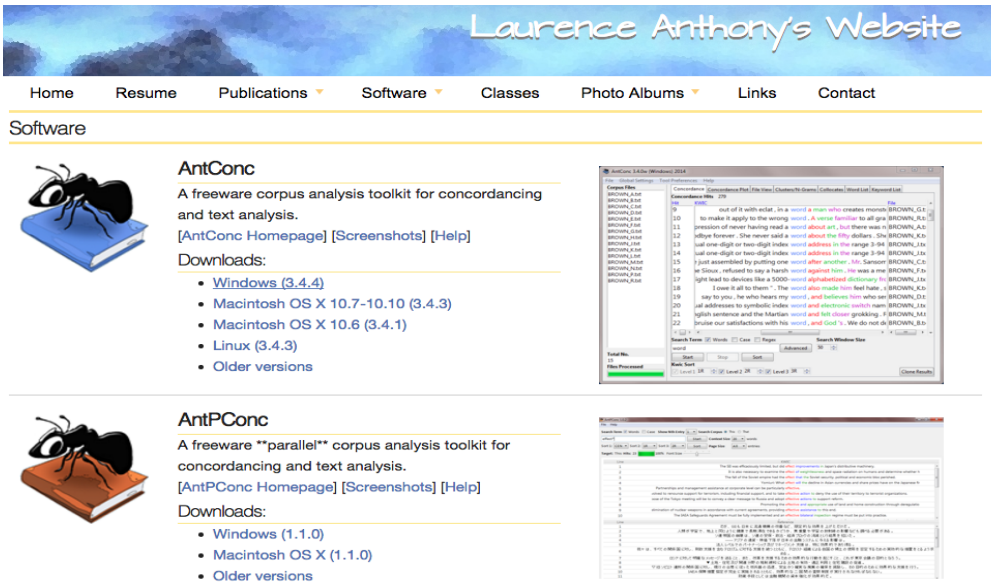
A guide to the use of AntConc in the sessions

We will use the AntConc software, version 3.4.4 and you can freely download from http://www.laurenceanthony.net/software.html. There are versions for Windows, Mac and Linux. Further information is available on the website.

1. Start up Internet browser and go to

http://www.laurenceanthony.net/software.html. you will see the following page.

Screen Shot 1*:* Loading AntConc from the website



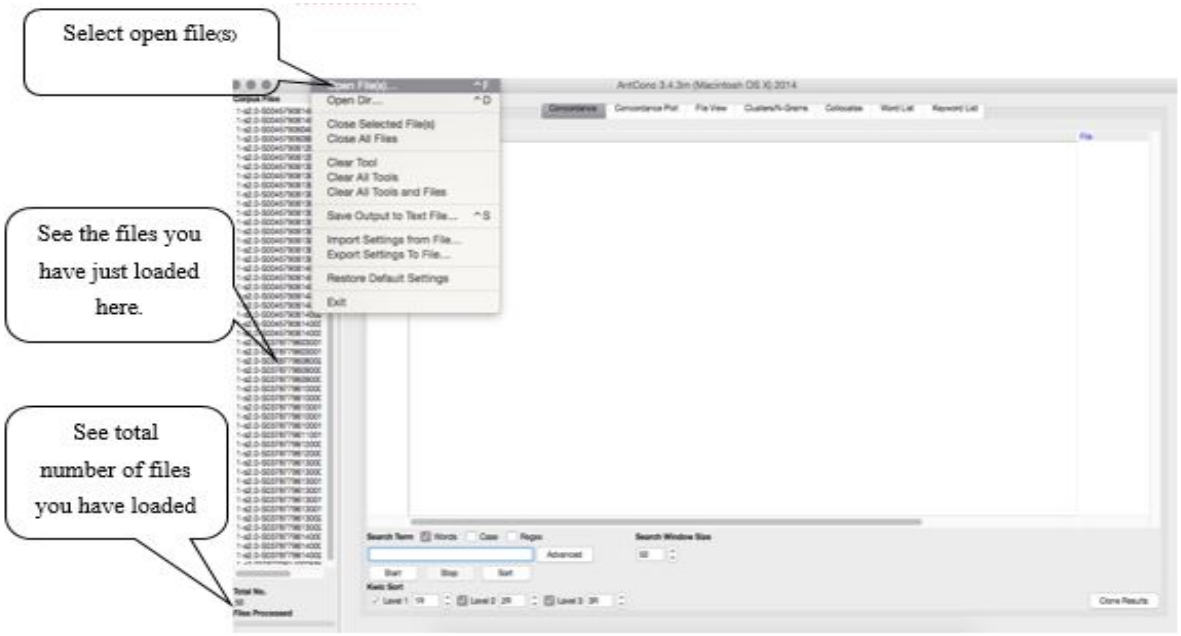
2. Click on Windows version 3.4.4, and you will be able to run the software. Now, you will be able to use the software.

3. On the file menu, select open file(s).

4. Select the folder where you have put your corpus and click OK. You should be able to have 50 txt. files (from computer and electrical txt. Files

Screen Shot 2*:* Loaded AntConc with a Corpus

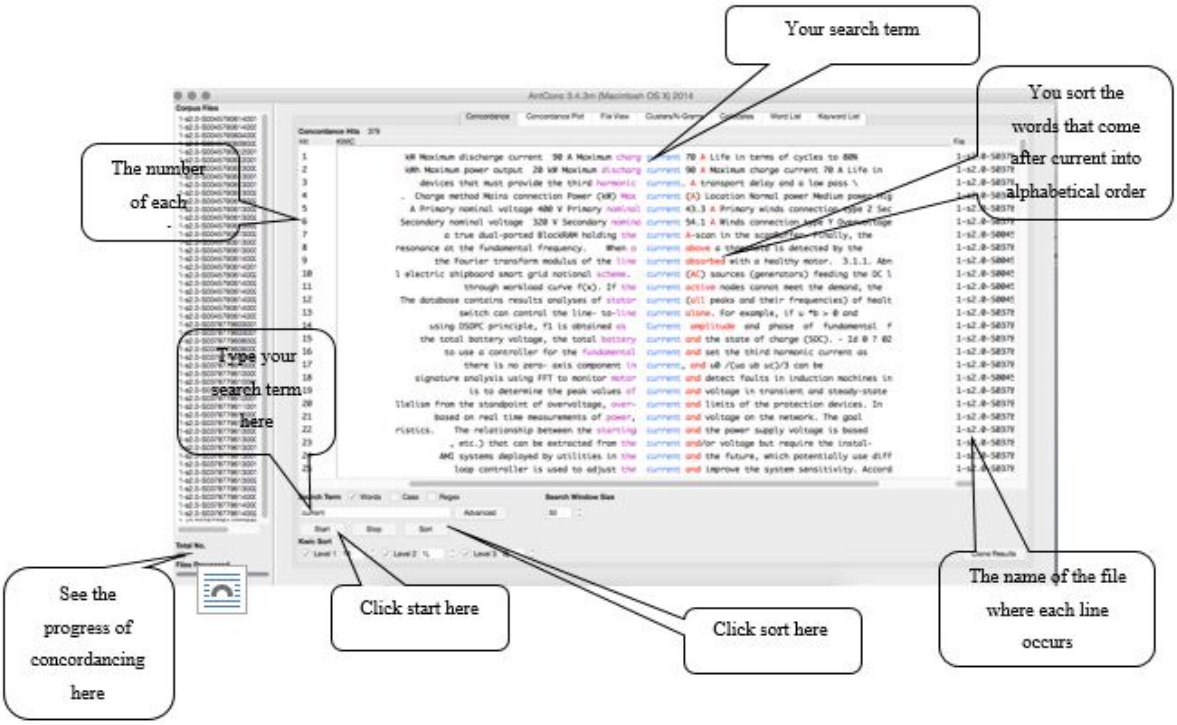
5. Type the word or phrase you want to find in the Search Term box. Type the word ‘current’.



6. Click the *start* button below the box.

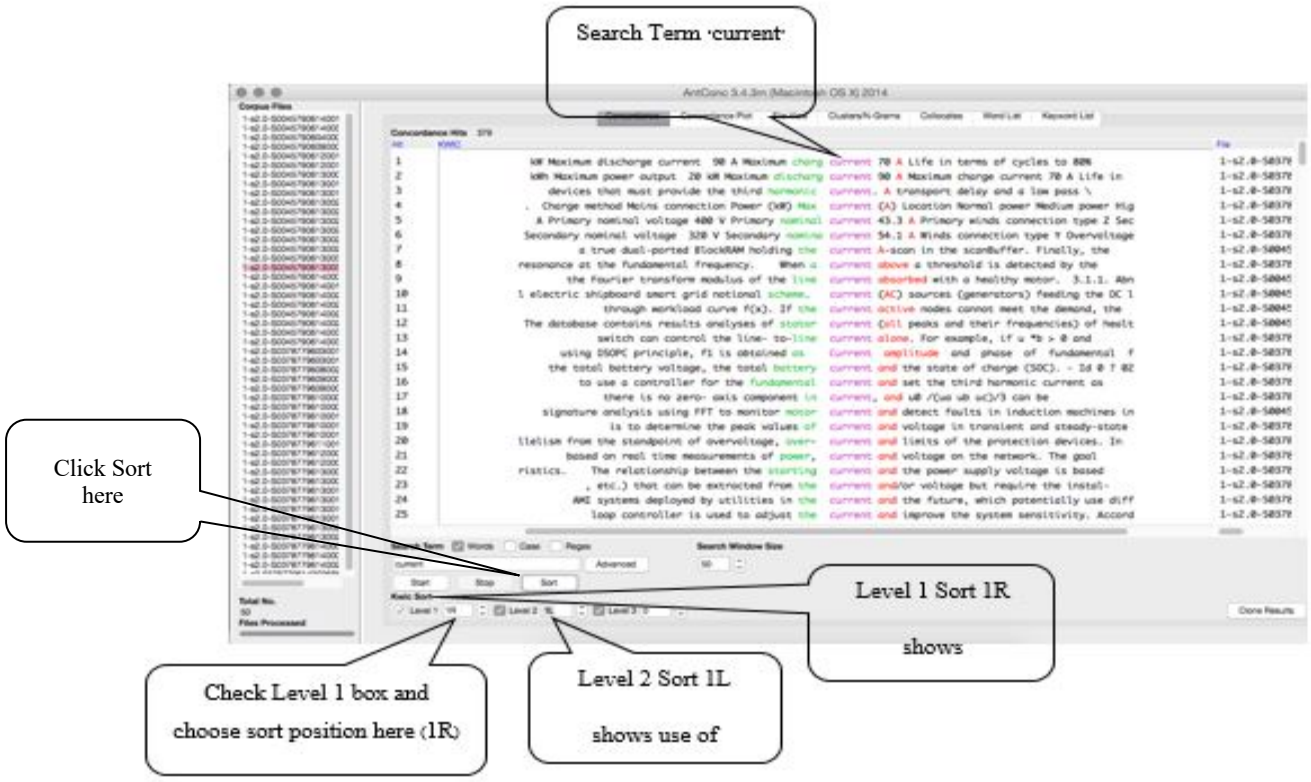
7. The concordance lines appear in the main window, with your search term highlighted in the centre.

Screen Shot 3*:* Concordance on the Word *'*current*'*



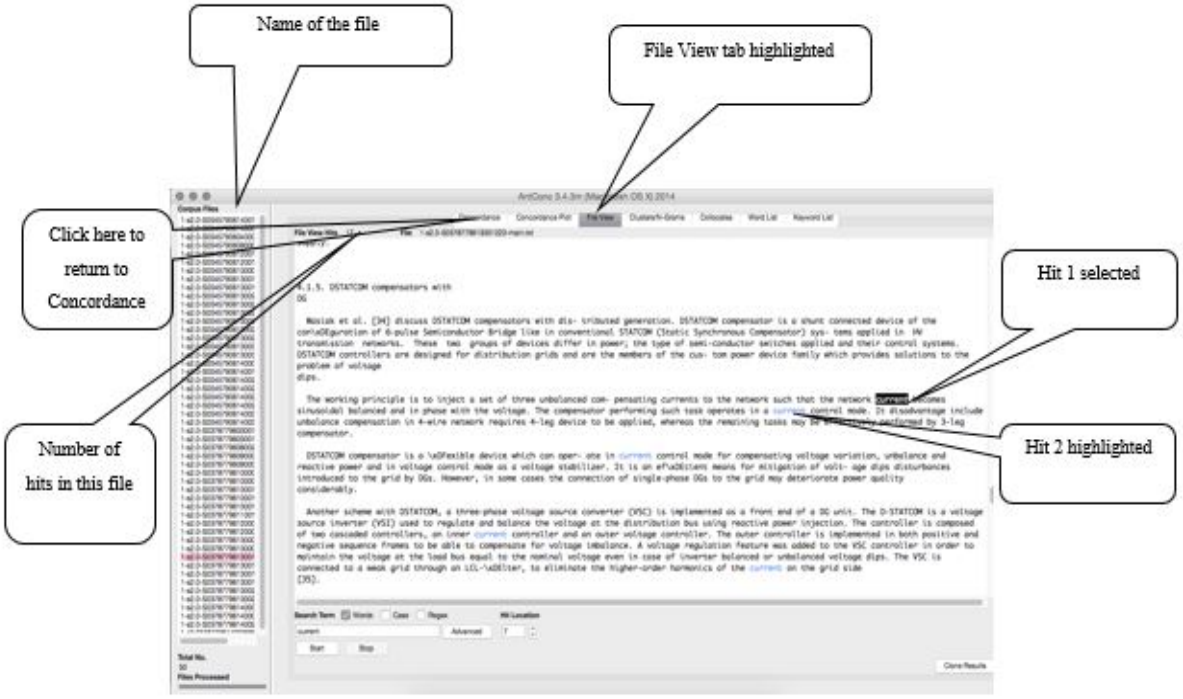
8. Click *sort* and you can sort the words that come before of after the target word into alphabetical order (from a to z).

Screen Shot 4*:* Concordance on the Word *'*current*'* Sorted by Level 1*:* 1R and Level 2*:* 1L



9. Change the basis of the concordance sort for example, 1R means one-word-to-the-right of *current,* 1L means one-word-to-the-left of *current,* and so on. For now, we want the 1R and then press *sort*.

Screen Shot 5*:* File view of the word *'current'*

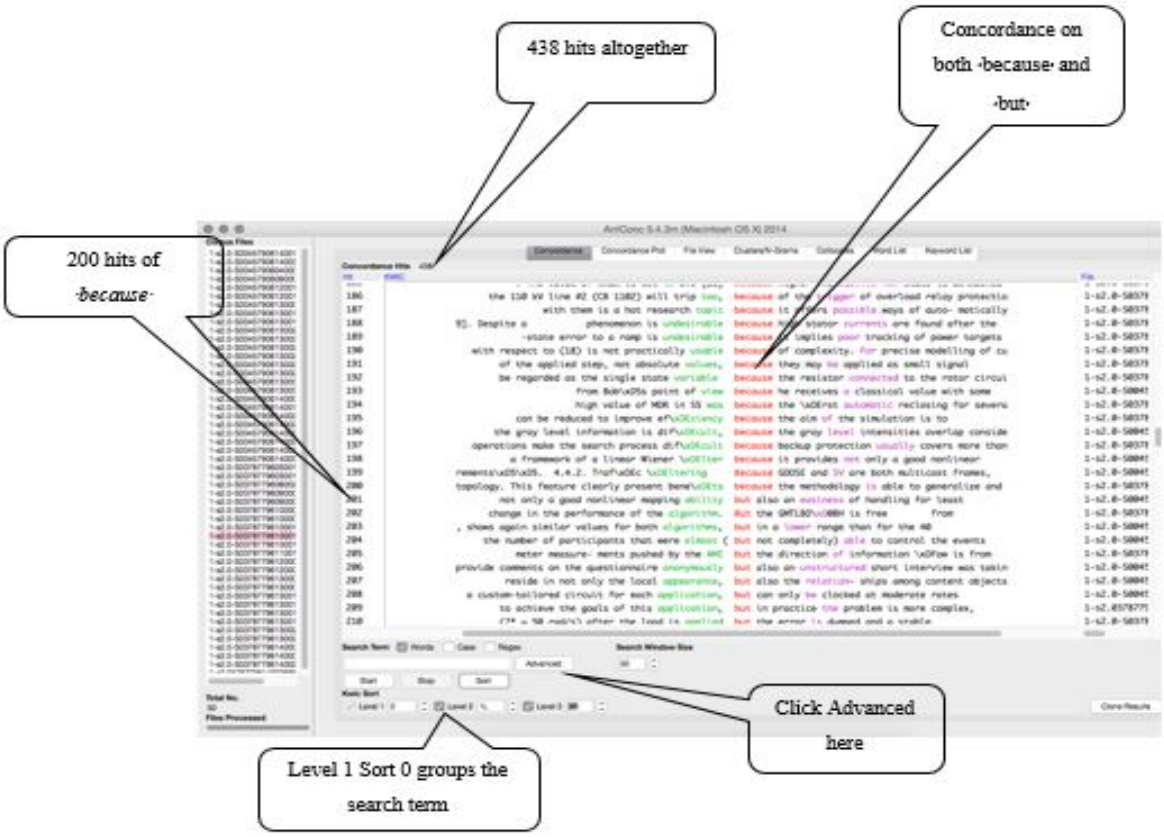


10. Click the search term, you will see the original file with the term '*current*' selected. It will bring you to the File View highlighted at the top of the window.

11. Look at the top of the window, it gives the total number of occurrences of your search term within that file and the name of the file. Here, the total number of the word *'current'* occurs in this file is 17 (hits).

12. Click the Concordance tab at the top of the window order to return to the concordance.

Screen Shot 6*:* Concordance on the words *'because'* and *'but'* at the same time



Screen Shot 7*:* Advanced Search Window for Search Terms *'because'* and *'but'*

13. Click Advanced Search to search for two or more terms at the same time.

14. Tick on the Use search term(s) from list below. Here, we want to search the terms '*because*' and '*but* 'at the same time.

15. Click 'apply' and this takes you to return to the concordance.

16. To sort the concordance, choose Level 1: 0 to group the lines.

17. You can go back to normal searching by unchecking the Advanced Search window.

Now, you are ready to work with the concordance and a corpus. Before you move on to activity 1, please use the uploaded 50 txt. files from electrical and computer as your main corpus.

Activity 1**:** Playing with AntConc and a corpus

Instructions: 1. After you have uploaded 50 txt. files (electrical and computer), study the word '*current'* from the concordance lines.

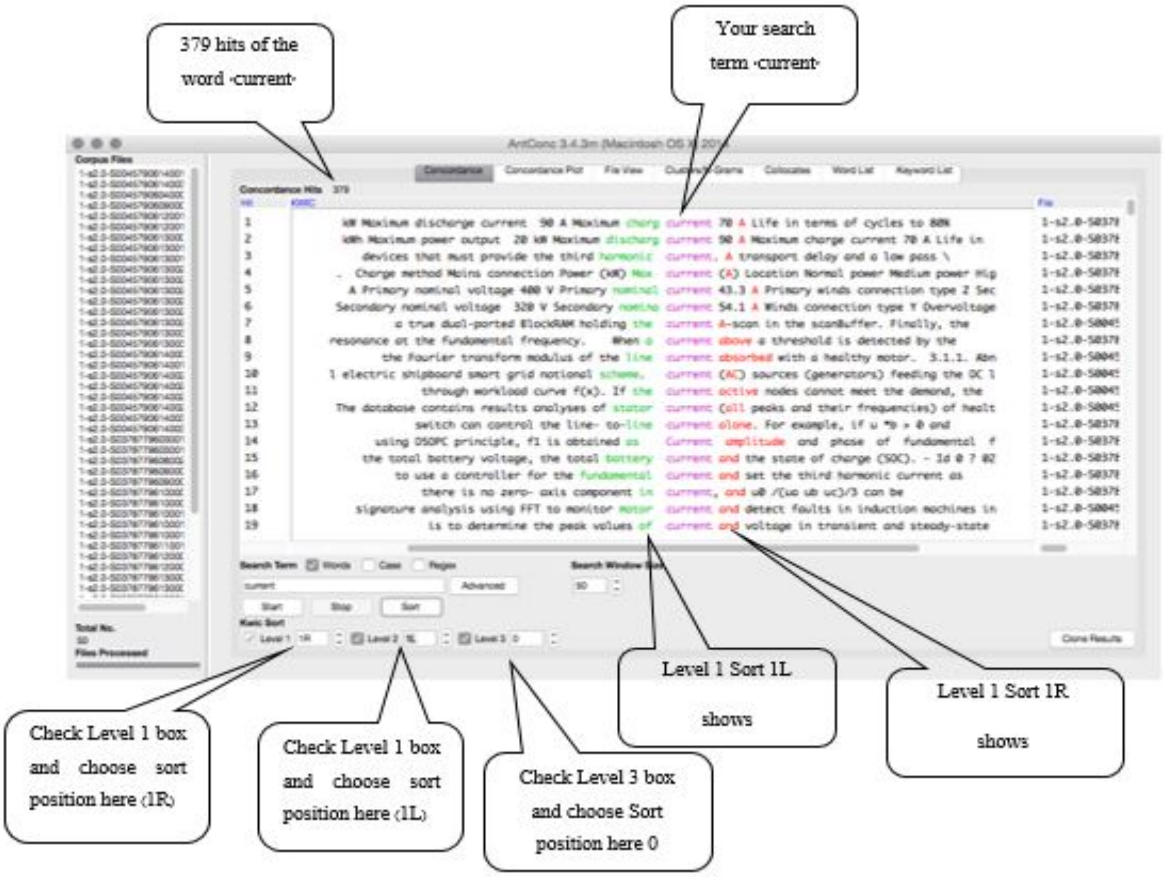
2. Notice the words that are used *before* or *after* the word '*current*' and try to list them in the same groups.

3. From the concordance lines, try to identify types of words (i.e.,nouns, verbs, adjectives, adverbs), and patterns of those words.

4. Share the possible answer with your friend near you.

5. After you have done with the uploading, you can compare your concordance results with the possible concordance results on the next page.

Screen Shot 8: Concordance on the word 'current'



Part 2: Training Students to Use AntConc and Analyse Lexical Bundles in ERAs Corpus

In Session 1, the training was divided into two main parts. The first part introduced the concept of corpora and concordance lines, focusing on helping students understand how corpus tools are used to analyse language data systematically. The second part involved hands-on training with AntConc software, where students explored lexical bundles in a small corpus of engineering research articles (ERAs). The primary aim of this session was to familiarise students with the functionality of AntConc and enable them to study lexical bundles' forms and functions within academic writing.

In Activity 2, for instance, students uploaded 50 text files (electrical and computer engineering-related) into AntConc. Using the software, they examined the word *current* in

concordance lines and noted the words that appeared before or after it. The task required students to group and categorise these surrounding words to understand their patterns, types, and roles in context.

S3's Experience with AntConc and the ERAs Corpus

This activity introduced students to the practical use of AntConc software and guided them in analysing the contextual use of *current*. Students were encouraged to explore the structure and function of the target word by identifying its surrounding words, their grammatical categories, and patterns. They also had the opportunity to work with additional bundles prepared for later activities.

S3's Response and Output:

S3 provided a detailed account of his experience with the activity, reflecting on its relevance and impact on his learning process. First, familiarity with the Software, S3 noted that the activity gave him confidence in navigating and using AntConc independently. He appreciated the user-friendly interface, which allowed him to examine words like *current* and their surrounding context efficiently. For instance, he commented, *"The tool made it easy to sort and group words based on their positions, such as 1R (one word to the right) and 1L (one word to the left). This feature helped me understand how the word current is typically used in academic contexts.* Second, understanding lexical bundles, S3 highlighted that the activity deepened his understanding of lexical bundles' structure and function. By analysing *current*, he discovered relationships with words such as *charge* (noun) and *absorbed* (past participle), which often appeared in its vicinity. He stated, *"I found it interesting to see how the words surrounding current fit into technical descriptions in engineering texts. It helped me understand how bundles form coherent academic expressions.* Third, preparation for subsequent activities, S3 explained that this initial activity provided a strong foundation for more complex analyses in later tasks. He remarked, *The hands-on experience with AntConc prepared me to study other bundles like the effect of and the number of. I now feel more confident analysing their structure and*

function in different contexts. Fourth, engagement and enjoyment, S3 reflected positively on the activity, describing it as both engaging and enjoyable. He stated, *I liked how interactive the task was. It broke down complex technical concepts into manageable steps, making it easier to follow and apply.* This level of engagement motivated him to explore the software further and build his competence in corpus-based analysis.

In terms of competence demonstrated, S3's reflections and outputs highlight his developing competence in several areas. First, technical proficiency: S3 demonstrated confidence in using AntConc to explore concordance lines and sort words by their positions, such as 1R and 1L. Second, linguistic analysis, he successfully identified patterns and types of words surrounding *current*, showing his ability to analyse lexical relationships within a corpus. Third, application of knowledge, S3 recognised the relevance of this activity for preparing him to analyse teacher-provided bundles in subsequent tasks. He understood how to apply these insights to explore more complex lexical patterns and academic writing conventions.

S3's experience with AntConc demonstrates the effectiveness of this activity in equipping students with the skills to analyse lexical bundles systematically. His ability to engage with the software, identify patterns, and apply these insights to his learning objectives reflects the success of the activity in fostering both technical and linguistic competence. By providing hands-on, interactive training, this session prepared students for more advanced tasks in lexical bundle analysis, aligning with the workshop's goals and addressing the reviewer's emphasis on demonstrating student competence.

> Assigned work **and Teacher's Emphasis on Lexical Bundle Structure and Functions**
> You have learnt the words from the concordance lines. Now, you should be able to notice the use of a single word in the concordance lines. You are ready to study groups of words. There is homework for you to practice more in the next activity. Please do the homework and share the answer with your friend next session.
>
> Instructions:
>
> 1. After you have uploaded 50 txt. files (electrical and computer), study the word '*the number of*' from the concordance lines.

2. Notice the words that use *before* or *after* the word '*the number of*' and try to list them in the same groups.
3. From the concordance lines, try to identify types of words (i.e.,nouns, verbs, adjectives, adverbs), and patterns of those words.
4. Share the possible answer with your instructor and your friends next time.

Activities in session 2

Session 2**:** Getting to know lexical bundles forms and functions in engineering research articles

In this session you will:

learn about the forms and functions of lexical bundles in the concordance lines taken from engineering research articles

be able to interpret the use of the lexical bundles (*the number of*, *the use of*, *the effect of*, *due to the*, *in order to*) in the concordance lines.

be able to identify patterns of the lexical bundles in the concordance lines

be able to produce a paragraph about your research you are working on by using the lexical bundles provided

Session 2**.**1**:** Lexical bundles forms in engineering research articles

Activity 1**:** where are the lexical bundles in the sentences?

Instructions:

1. From the nine sentences below, underline *the use of, the effect of, the number of*.
2. From the provided sentences, notice how each lexical bundle (*the use of, the effect of, the number of*) is placed and used in the sentences, such as in the beginning or the middle of the sentence, as a subject, or as an object.

1. The use of a local product depends on its specific properties and the requirements for a particular job.
2. If their claim is valid, then adequate protection of concrete should be ensured by the use of a low w/c alone or by the use of Type V cement alone.
3. The integration has been possible thanks to the use of a customized communication protocol, whose details of implementation are discussed. Finally Section 5 reports the conclusions.
4. It is instructive to see the effect of additive on creep and relaxation behavior.
5. The effect of Ca2+ intrusion reflects itself in the different gradings of SAPD.

6. A subroutine was written to consider the effect of conduction, convection and radiation during the laser deposition process

7. We analyse the influence of (a, b) pairs on MSA features as well as a range of suitable values, while pointing out the relationship between the number of (a, b) pairs and the recognition rate.

8. The number of added FFs should be large enough so that there is a high probability that PUF response sets the initial power-up state to one of the added states.

9. Where N is the incremental frequency (the number of AE hits with amplitude greater than the threshold), M is earthquake magnitude or Richter magnitude of events, a is an empirical constant and b is the AE based b-value.

Analysis of Activity: Assigned Work and Teacher's Emphasis on Lexical Bundle Structure and Functions

Overview of Activity: This activity builds upon students' initial understanding of concordance lines and introduces them to the analysis of lexical bundles in engineering research articles (ERAs). The focus is on studying the lexical bundle *the number of* in context, identifying surrounding words and patterns, and categorising these into grammatical types (e.g., nouns, verbs, adjectives). Students were tasked with analysing concordance lines to deepen their understanding of lexical bundle forms and functions. The teacher emphasised the importance of understanding the structure and function of lexical bundles, preparing students to produce their own writing using these bundles.

S3's Experience and Reflections:

S3's account provides valuable insights into how this activity facilitated his learning process and demonstrates his engagement and competence. First, understanding the structure of lexical Bundles, S3 reflected on how analysing *the number of* in concordance lines helped him grasp its structure and function within academic writing. He stated, *I noticed that the words surrounding 'the number of' were often nouns or verbs, which made me think about how this bundle is used to quantify or describe something in technical contexts.* For example, he identified words such as *added* (verb) and *FFs* (noun) following *the number of* and recognised their role in conveying technical details.

In terms of competence categorisation of words and patterns, S3 demonstrated competence in categorising the surrounding words based on their grammatical roles. He explained, *By using AntConc, I could sort words that came before and after the bundle and group them into types like nouns and verbs. This made me understand the patterns more clearly.* His ability to categorise these words shows his growing understanding of how lexical bundles fit into larger linguistic structures.

Regarding preparation for writing and subsequent tasks, S3 highlighted how this activity prepared him for subsequent tasks, such as producing paragraphs using lexical bundles. He mentioned, *After this activity, I felt more confident in identifying bundles in sentences and thinking about how to use them in my own research writing.* This demonstrates the activity's success in bridging analysis and application.

This activity enhanced engagement and teacher support**.** S3 appreciated the clear instructions and the teacher's scaffolding during the activity. He noted, The teacher's explanation about how *the number of* is used in different contexts really helped. It was also useful to discuss my findings with classmates. This collaboration and guidance further enhanced his understanding and engagement.

Moreover, the activity might have challenges when doing the activities to and at the end S3 can overcome them. While S3 found the activity engaging, he acknowledged some initial difficulties in understanding the grammatical patterns of bundles like *the number of*. He stated, At first, it was hard to figure out how the surrounding words worked with the bundle, but after practising and seeing more examples, it became clearer. This suggests that repeated exposure and teacher feedback were critical in overcoming these challenges.

S3 demonstrated developing competence in linguistic analysis by identifying and categorising words surrounding the bundle *the number of* and recognising its patterns in academic contexts. He effectively applied theoretical knowledge by connecting concordance line analysis to his writing. Active participation in peer discussions and teacher-guided activities further enhanced his understanding. This activity successfully introduced students to analysing lexical bundles, focusing on their structure and function. S3's reflections illustrate the activity's role in building competence with concordance tools

and analysing lexical patterns. By showcasing detailed examples of his learning, this analysis addresses the reviewer's comment on demonstrating student understanding.

Self-study section: Subject and Object

Subject

The subject is a noun or a pronoun that comes before the verb in an ordinary affirmative sentence. An affirmative sentence is often called an active sentence and the subject is who or what does the action that the verb refers to (Adapted from Swan, 1996).

For example, *the engineer* is doing the experiment. This sentence is called an active sentence, and the engineer is the subject and he does the action (experiment).

Object

The object can be also a noun or pronoun that normally comes after the verb, in an active clause. The direct object refers to a person or thing affected by the action of the verb, for example Take *the dog* for a walk (The dog is a direct object of this sentence). The indirect object usually refers to a person who receives the direct object e.g.,Ann gave *me a watch* (me is an indirect object and the direct object is a watch) (Adapted from Swan, 1996).

Activity 2:

Instructions: On your own, look at the words in bold and decide whether they are the subject (S) or the object (O). Compare the answer with your friends.

1. Radial basis function (RBFs) neural networks as a kind of powerful kernel methods have been applied to many areas with success.
2. The theoretical analysis of RBF structures and algorithms includes the orthogonal least square algorithm, the approximation capability analysis [17,18,20], the design of RBF structure using fuzzy clustering method, the optimization of RBF structure using kernel orthonormalization method or combined supervised and unsupervised learning method, and the use of Fisher separability ratio for the selection of RBF centres.
3. The RBF is selected because of its compact support [17,18,20].
4. Regarding Wiener, it has been implemented with different structures for noise suppression applications as MEG signals, acoustic signals, etc. [21].
5. The effect of percentage of VPA on the permeability of VPC at the age of 28 days is shown in Fig. 2.

Now, you know what the subject and the object in the sentences. So, try to review them again by doing Activity 3 below. You can discuss with your friends near you, if necessary.

Activity 3

Instructions: Can you underline the subject and the object of the following sentences.

1. The permeability is increased from approximately 3.6 x 10 - 10 cm/s to around 13 x 10 - 10 cm/s when VPA content is increased from 0% to 100% by volume.

2. Compared to normal (0% VPA) concrete, the permeability of VPC with 100% VPA is about 3.5 times greater.

3. The changes in the number of equilibria now take place at higher values of the load.

4. The poorly damped oscillatory behaviour has been avoided producing a smoother operation.

5. In Eq. (9) the amount of lubricant delivered by the lubricator is calculated by counting the number of lubricant drops delivered in a given time period (usually 120 s).

What have you noticed from the above activities? We can say that the subject can be a single noun and can include all modifiers that go with it. And the object can be noticed by …?.

Homework!

Instructions:

1. Study the group of words *the effect of* from the concordance lines.

2. Notice the words that use *before* or *after* the word *the effect of* and try to list them in the same groups.

3. From the concordance lines, try to identify types of words (i.e.,nouns, verbs, adjectives, adverbs), and patterns of those words.

4. Share the possible answer with your instructor and your friends next time.

Activity related to identification of the patterns of the lexical bundles

Instructions:

1. The following patterns tell us about how the lexical bundles (*the effect of, the number of*) are used in the research articles. Can you group the right concordances with the appropriate pattern?

2. When finished, you can share the answer with your neighbours and then you will be given the possible answer together as a whole class.

3. Can you guess which section of the research article each sentence comes from?

1. However, because the quantity of chloride ions absorbed by C–S–H gel is far lower than that bound by FS [42], *the effect of* decomposition of C–S–H gel on the stability of bound chlorides is insignificant when compared with the decomposition of FS under MgSO4 attack.

2. However, the values of radial stress at the inner and outer boundaries of the cylinder are not zero because of the effect of prescribed pressure loadings.

3. When we add a large number of new states |S0|, |S| + |S0| states can be implemented by a linear growth *in the number of* FFs that is log (|S| + |S0|).

4. Actually, when the dimensions of the generator are reduced, the induced voltage is decrease inherently as a result of decrease in the total flux even if *the number of* conductor remains the same.

5. Müllauer et al. investigated the effect of external sulfate attack on the leaching behavior of heavy metals in concrete.

6. Fig. 6 shows *the effect of* pile spacing on the thermally-induced mechanical behavior of energy pile (3*3 arrangement) in a group in sand.

7. This study has enabled investigating *the effect of* the water distribution system on the thermal performance of a forced draft counter-flow cooling tower (FDCT) filled with six different types of drift eliminators.

8. The only difference is that they were immersed in a similar covered container of 5% MgSO4 solution for 28, 56 and 90 days at the standard curing condition (20 ± 2 °C and 95% RH) for the examination of *the effect of* MgSO4 attack on the stability of bound chlorides.

9. Hajidavalloo et al. [18], in their work of cross flow cooling towers in variable wet-bulb temperature, include a brief discussion about *the effect of* the drift eliminator on tower performance, taking only a reduction of airflow rate into account.

10. The parameter is attributed in such a way that the total number of generated clones is twice *the number of* B cells in the population, i.e., nc = 200.

11. The effect of frequency and power was characterised on the depth and efficiency of concrete removal.

12. A reduction in *the number of* buffers improves the power efficiency; whereas, keeping a few necessary buffers intact, improves the performance.

13. In order to evaluate *the effect of* thermoelectric leg geometries on the power generation performance of thermoelectric modules, finite-element thermoelectric analyses were performed on rectangular-, trapezoidal-, octagonal-prism, and cylindrical models for the temperature gradient of 100 ºC.

14. *The number of* added FFs should be large enough so that there is a high probability that PUF response sets the initial power-up state to one of the added states.

15. *The number of* nodes can be reduced to improve efficiency because the aim of the simulation is to explore the relationship between the TF variations and the severity of the inter-turn fault rather than build a detailed numerical model for a specified winding.

Try doing the following exercise, assigning each of the sentences above to one of these patterns. BUT, if you are not sure what they are, there is an explanatory section about *clause* and *phrase* in the *self-study section* provided for you.

Pattern 1: beginning of a clause or phrase + *the effect of* or *the number of* as a dependent complement clause or prepositional phrase

Pattern 2: beginning of a clause with *the effect of* or *the number of* embedded + a prepositional phrase

Pattern 3: begins with *the effect of* clause + a prepositional phrase

Pattern 4: begins with a phrase embedded with *the effect of* or *the number of* clause plus a dependent complement clause or a prepositional phrase

Self-study section: Reviewing phrase and clause

A *phrase* consists of two or more words that function together as a group, such as a large long machine, in this study, the effect of percentage of VPA on the permeability of VPC at the age of 28 days.

In contrast, a *clause* contains a subject and verb, and it can convey a complete idea. The clause is joined to the rest of the sentence by a conjunction e.g.,*Mary said* that *she was tired*. There are two types of clause: an independent clause and dependent clause.

An independent clause can express a complete thought (and can be a standalone sentence). A dependent clause is usually a supporting part of a sentence, and it cannot stand by itself as a meaningful proposition (idea).

The clause is sometimes contained participles or infinitives (with no subject or conjunction), such as *Not knowing what to do*, I telephoned Robin; I persuaded her *to try a new method.* You can see that the first sentence, '*Not knowing what to do*, I telephoned Robin.' contains present participle ending with '-ing' form (Not knowing what to do). The infinitive 'to + base verb' in the second one is 'to try a new method'. (Adapted from Swan, 1996)

Activity 1:

Instructions: On your own, identify independent and dependent clauses in the following sentences. Check the answer together in class.

1. Before beginning each rheological measurement, the mortar was agitated using a spoon in order to reduce the effect of possible sedimentation.
2. Thirdly, the use of the term "Na2O-equivalent" implies that the effect of potassium and sodium is equivalent.

3. However, there are numerous studies indicating that the influence of potassium and sodium on silica and quartz dissolution is not identical [e.g.,[18,19]].

4. Despite the work on the interaction of sulfate and chloride in concrete mentioned above, there is very little work on the effect of sulfate attack on the stability of bound chlorides in concrete.

5. In order to better capture the effect of the segment on the efficiency, the TE calculation is carried out at a constant load resistance condition.

Activity analysis: subject, object, and lexical bundle patterns

Understanding the subject and object in sentences enables students, including S3, to identify lexical bundle patterns in engineering texts. This set of activities was designed to develop students' skills in recognising subject-object relationships and the structural patterns of lexical bundles, such as *the effect of* and *the number of*. Students followed a step-by-step process involving self-study, identification tasks, and peer discussions, culminating in homework that encouraged independent analysis.

Regarding S3's experience and competence demonstration, S3 highlighted that working through the subject-object identification exercises deepened his understanding of sentence structure in academic writing. He noted that identifying the subject and object helped him see how lexical bundles fit into sentences and connect with surrounding words. For example, in sentences like *the effect of percentage of VPA on the permeability of VPC is shown in Fig. 2*, S3 successfully identified *the effect of* as part of the subject and recognised how it related to the predicate. While working on the lexical bundle analysis, S3 engaged with concordance lines to categorise surrounding words into grammatical types (e.g., nouns, verbs) and patterns. He found it interesting to see how words like *chlorides* (noun) and *absorbed* (verb) often followed *the effect of*. This helped him understand how bundles function to describe technical processes. His ability to group similar patterns and understand their roles in academic contexts demonstrated his growing competence.

S3 acknowledged some initial difficulties in distinguishing dependent and independent clauses, especially in longer sentences. He explained that it was confusing at first to figure out which part was the clause and which was the phrase. However, the self-study materials and teacher feedback provided much-needed clarity, enabling him to proceed with more confidence.

S3 demonstrated competence in linguistic analysis by successfully identifying and categorising subject-object relationships and lexical bundle patterns in sentences, which enhanced his grammatical understanding. He applied this knowledge by analysing concordance lines and connecting his findings to the practical use of bundles in academic writing, showcasing his ability to bridge theoretical concepts with real-world application. Additionally, S3 actively participated in peer discussions, which allowed him to gain diverse perspectives and reinforce his understanding of the material. These experiences highlight his growing proficiency in using corpus tools and analysing lexical bundles, reflecting the effectiveness of the structured activities and support

Knowing how to identify the subject and object in sentences is essential for understanding lexical bundle patterns in engineering texts. This activity effectively guided students through these analyses. S3's reflections show how the structured tasks and teacher support helped him build competence in linguistic analysis and apply these skills to academic contexts. By including detailed examples of his learning process, this analysis addresses the reviewer's comment and highlights the activity's value in fostering student engagement and understanding.

Activity C: Knowing about engineering actions by the use of lexical bundles

Instructions:

1. Notice the use of the lexical bundles (*the use of, the effect of, the number of*) in the following concordance lines.
2. Underline the words that follow the lexical bundles i.e., *the effect of, the use of, and the number of*. Identify what types of word usually follow them such as, nouns, verbs, collocations, adverbs. Then, complete Table 1 below.
3. Do you think the words that follow the three groups of words play an important role in the sentence, and indicate specific actions in engineering? How?

1. The risk may then be compared with a standard defined by the utility or by a regulatory authority, in order to check whether or not it is necessary to increase *the number of* committed units in that situation.
2. *The effect of* percentage of VPA on the permeability of VPC at the age of 28 days is shown in Fig. 2.

3. However, *the use of* a high end computer and the wired installation per home increases the expense of the system.

4. *The use of* a multi-physic software plays a fundamental role for the design of power systems, due to that great part of the challenges are not limited to electrical and communication network but also involve other domain and energy networks (e.g.,heating and thermal) [7]

5. Ref. [12] involves *the use of* a networking cloud, Pachube, which is a real time data infrastructure that allows the management of data points from individuals, organizations and companies through the Internet.

6. The voltage divider model was used to study *the effect of* DG units on voltage dip.

7. Although there are detailed differences in the temperature distributions because of *the effect of* different convective cooling and the different disc geometry, overall the results demonstrated that the scaling methodology can be used with confidence for the design and development of automotive disc brake systems.

8. A reduction in the number of buffers improves the power efficiency; whereas, keeping a few necessary buffers intact, improves the performance.

9. This is because *the number of* alternative paths that a header can take to progress changes as it advances towards the destination.

10. If *the number of* faulty links surpasses the available spare links, then the split transmission and packet re-organization are performed.

Table 5 Answer for Activity C

Item	Groups of words	Words that co-occur	Types of words	Indicate what
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				

Activity C: Understanding engineering actions through lexical bundles

This activity aimed to help students analyse lexical bundles (the use of, the effect of, the number of) within engineering contexts. S3's engagement with this activity highlighted his developing competence in analysing concordance lines and identifying patterns of lexical bundles. S3 effectively identified words that followed the bundles, such as percentage (noun) in the effect of percentage of VPA on the permeability, and multi-physic software (noun phrase) in the use of a multi-physic software plays a fundamental role. He categorised these co-occurring words into grammatical types (e.g., nouns, prepositions, adjectives) and recognised their role in defining engineering actions. For instance, S3 noted that the effect of often introduces a description of influencing factors, while the use of highlights tools or methods in technical processes. Similarly, the number of frequently quantifies or specifies variables, as seen in the number of faulty links surpasses the available spare links. These observations demonstrate his ability to connect grammatical analysis to the practical functions of bundles in engineering texts.

In terms of competence student (S3) gained competence in several key areas during this activity. Through linguistic analysis, he identified and categorised co-occurring words, such as nouns and prepositions, showcasing his understanding of lexical bundle patterns. By analysing concordance lines, S3 successfully linked the grammatical structure of bundles to their functional roles, such as describing tools, actions, or influencing factors in engineering texts, illustrating his ability to apply knowledge effectively. Additionally, his critical thinking skills were evident as he evaluated how lexical bundles contribute to sentence meaning and signal technical processes or problems, demonstrating a deeper understanding of their role in academic writing. S3’s engagement and reflections provide concrete evidence of his understanding and competence. His ability to classify and interpret lexical bundle patterns shows the effectiveness of the activity in enhancing linguistic and analytical skills. By presenting examples of S3's responses, such as his grammatical classifications and insights into the roles of bundles, this activity directly addresses the reviewer’s comment by demonstrating student learning outcomes and engagement.

In conclusion, the activities in this session helped students improve their understanding and use of lexical bundles in engineering research articles (ERAs). The tasks were designed to guide students step by step, starting with basic grammar and moving to more complex analysis of concordance lines. Student reflections, like those from S2 and S3, show how these activities supported their learning.

S2 showed progress in understanding word meanings in context and recognising patterns, which helped her prepare for more advanced tasks. S3 demonstrated growth by identifying subject-object relationships and grouping co-occurring words, improving his understanding of how bundles work in sentences. Both students appreciated the clear instructions and opportunities to discuss their ideas with peers, which helped them gain confidence and overcome challenges.

The activities combined different skills like analysing language, applying knowledge to real examples, and thinking critically about how bundles are used in sentences. Peer discussions and teacher support played a key role in strengthening their learning. These tasks not only helped students use lexical bundles in academic writing but also taught them how to evaluate grammar and technical content in engineering texts.

The students' responses and work clearly show how the activities improved their skills, addressing the reviewer's comment by providing evidence of their learning. The structured design of the tasks, along with examples and reflections, highlights their effectiveness in preparing students for academic and professional success in engineering communication.

4.2.2.2. Piloting the materials

The aim of piloting and trialling is to check the validity of the materials, mainly checking whether the materials match the purposes of the activities. There were 11 participants in the pilot and trial: five lecturers from the English Department, two lecturers from the Engineering Department, two lecturers from the Social Sciences Department and two engineering students. Pilot participants were asked to comment on four main points: clarity

of the questions in the materials, language used in the materials, signposting of the materials, activities order.

4.2.2.3. Activities

In this study, the term 'materials' refers to the instructional resources specifically developed or selected to facilitate the learning of lexical bundles in engineering contexts. Examples of materials include worksheets, handouts with authentic lexical bundles, and examples drawn from engineering research articles. These materials are designed to provide learners with meaningful exposure to and practice in using lexical bundles. The term *activities* refers to the classroom tasks or exercises used to engage students in the learning process actively. These activities include analysing lexical bundles in authentic texts, group discussions about their functions, collaborative creation of lexical bundles for specific engineering scenarios, and peer evaluations of those creations. The purpose of these activities is to promote not only linguistic accuracy but also critical thinking skills by requiring students to apply and reflect on their knowledge in practical contexts.

Nine out of 11 participants thought that the ways the activities were organized seemed to be satisfactory. For example, one of them agreed that the logic of the activities can promote *analysing skills*. Analysing skills in this study refers to the ability of students to critically evaluate and interpret the structure, function, and usage of lexical bundles in authentic texts. These skills are developed through activities such as examining lexical bundles in context, identifying their patterns, and discussing their relevance to engineering-specific scenarios. The self-study section was quite helpful in reinforcing language points for the students, particularly regarding subjects and objects activities.

Regarding the identification of *subjects and objects* activities, they were designed to help students understand and analyse the relationship between subjects and objects in sentences, as well as recognise the structural patterns of lexical bundles in engineering texts. These activities included multiple steps to engage students and develop their analytical skills. First, students completed a self-study section where they reviewed materials explaining the concepts of subjects and objects in sentences. This section included examples of lexical bundles, such as *the effect of* and *the number of*, to illustrate

how subjects and objects interact with lexical bundles. Next, students participated in identification tasks, where they analysed engineering-related texts to locate sentences with lexical bundles and identify the subject and object in each sentence. This helped them understand how the subject-object relationship impacts the meaning and function of lexical bundles. Following the identification tasks, students engaged in peer discussions to share and compare their analyses. These discussions encouraged collaborative learning and allowed students to deepen their understanding of how lexical bundles function in context. Students then applied their learning by creating their own sentences using specific lexical bundles, ensuring accurate use of subjects and objects within an engineering context. Finally, as part of their homework, students completed independent analysis tasks to further reinforce their skills by identifying subjects, objects, and lexical bundles in new sentences or short texts. These activities were carefully designed to provide a step-by-step approach, allowing students to gradually build their skills in recognizing and applying subject-object relationships and lexical bundle patterns in meaningful contexts. The combination of guided practice, collaboration, and independent work ensured that students could fully engage with and understand these linguistic features.

This activity is designed to increase students' awareness of how subjects and verbs function within texts, particularly in the context of engineering-related lexical bundles. By engaging in self-study, identification tasks, peer discussions, and independent analysis, students develop their ability to recognise and analyse subject-object relationships and the structural patterns of lexical bundles. While the activity effectively promotes linguistic awareness and critical thinking, there are still considerations to address in refining the materials further to create ideal materials tailored specifically for the engineering context.

4.2.3. Issues that need to be considered

4.2.3.1. Theory of teaching

One of the participants raised an important question: should teaching theory be involved in the materials? This study was conducted as a preliminary stage of integrating a lexical approach and using corpus data, referred to as Engineering Research Articles

(ERAs), with engineering students. Given this focus, the primary aim was to explore how students engage with and analyse the structure and patterns of lexical bundles independently. The study has already incorporated elements of the Presentation, Practice, and Production (PPP) teaching model through activities that encourage students to discover the structure and function of lexical bundles on their own. These activities emphasise independent discovery and analysis rather than a detailed implementation of specific teaching theories. As the use of teaching theory is not the focus of the study, its explicit integration into the materials may not be necessary at this stage. The study prioritises the development of materials that facilitate student engagement with lexical bundles through hands-on, discovery-based tasks rather than a structured application of teaching principles. While teaching theory could play a role in future stages of research, the current study focuses on the practical application of corpus-based methods to enhance students' linguistic awareness and analytical skills.

4.2.3.2. Language use in the classroom

One of the participants expressed concern that students might experience difficulty when studying this material in class, particularly during tasks that involve identifying and categorizing lexical bundles in unfamiliar texts. Specifically, the participant noted that students may find it challenging to recognise the structural patterns and communicative functions of lexical bundles—such as distinguishing between noun phrase bundles and prepositional phrase bundles—especially if they are unfamiliar with analysing linguistic features in authentic academic texts. However, the participant also suggested that with appropriate guidance and support, such as providing clear instructions, step-by-step examples, and scaffolded tasks (e.g., matching bundles to functions before identifying them independently), students would be able to complete the exercises successfully. Arranging activities in a progressive manner—starting with familiar examples and gradually moving to more complex and discipline-specific texts—was highlighted as a key factor in helping students overcome these difficulties.

It was suggested that certain terms and questions in the activities should be revised for better clarity and to reduce confusion among students. For instance, the word *notice* in

Activity C could be replaced with phrases like *pay attention to* or *study'* However, I believe the word *notice* encourages students to explore lexical bundles independently, even though it might cause some confusion with certain questions and terms. For example, I should clarify what is meant by *target word* to avoid ambiguity. Similarly, the phrase *potential meaning* in Activity 2 could confuse students, so I should consider rephrasing it as a question, such as *Is important commonly used in positive or negative contexts?* Additionally, the questions *What is the verb that is always used before important?* and *What other types of words are used before important?'*may hinder students' understanding, particularly due to the use of the phrase *types of word*. Engineering students seemed to struggle with understanding what *types of words* means, so this terminology should be simplified or explained more clearly.

4.2.3.3. Activities

One of the participants, a PhD engineering student, found the materials to be highly beneficial in improving his research article writing skills. However, he suggested that the activities in Part 1, which focus on using AntConc, should be revised to include more detailed instructions. He expressed concern that students might lack confidence if they are unable to effectively use the program. Drawing from his perspective as both a teacher and a student, he provided several valuable suggestions for improving the activities, aiming to make them clearer and more accessible for learners.

The participant provided several valuable suggestions for improving the activities to better support students in using AntConc and engaging with the materials effectively.

First, the terms *1L* (one-word-to-the-left) and *1R* (one-word-to-the-right) should be clearly explained, specifying that students can search only *1L*, only *1R*, or both together. Additionally, the *KWIC sort* section should include an introduction to levels 1, 2, and 3 to help students understand how to use this feature. It was also recommended to clarify ambiguous phrases, such as *it looks okay* on page 6, to ensure students know when they are ready to proceed with a search using the concordance tool. The instructions in Section 1 should be written in an affirmative form and include a step-by-step explanation to help students follow along more easily and avoid feelings of technophobia, which could arise if

they struggle with the program. Furthermore, the activities should be tailored to specific engineering disciplines by using lexical bundles frequently found in those disciplines, allowing students to write paragraphs relevant to their own field. Alternatively, foundational subjects common to all students, such as basic engineering, mechanics, and physics, could be used initially, with students later working in groups on discipline-specific activities. This approach ensures students are engaging with content they are familiar with, which can boost their confidence and improve their writing in their specific fields. Finally, the concordances selected for teaching should be limited to no more than 10 lines to accommodate the students' level of proficiency. This does not mean that students are unable to read longer texts; rather, it reflects the need to introduce concordances in a manageable way when students are first working with them. A limit of 10 lines is appropriate for beginners as it aligns with their current skills and allows them to focus on analysing the patterns and structures of lexical bundles without becoming overwhelmed.

4.2.3.4. Sequence of the activities

Based on student feedback, it was suggested that the sequence of activities be rearranged to introduce concordances and node words before demonstrating how to use AntConc. This adjustment could make the initial introduction to Data-Driven Learning (DDL) more accessible to students, particularly those unfamiliar with corpus tools. Additionally, providing handouts of pre-printed and simplified concordance lines as a preliminary activity would help students understand the concept of concordances and focus on recognizing patterns without the added complexity of navigating software. This approach aligns with best practices in introducing DDL and ensures that students build confidence before engaging with more advanced tasks. These suggestions have been incorporated into the recommendations for improving the instructional design.

In the section *searching for a word and seeing its concordances*, it was recommended that I should not start with the phrase *the number of* because it might be too hard for students to understand at first. One of the pilot testers suggested starting with something simpler, such as asking students to explore the word *house*. This approach would help introduce the concept of a node and explain what can be learned from concordance

lines. It would also allow me to demonstrate effective techniques for exploring concordances. Following this explanation, I could give students a few minutes to practise using the tool by searching for words of their choice.

4.2.3.5. Clarity of the instructions*:* the instructions should be clearer and more specific

Some improvements were suggested for the instructions and content of activities B and C to make them clearer and more accessible for engineering students. For instance, the instructions in activity B, *The following patterns tell us about how lexical bundles (the effect of, the number of) are used in research articles. Can you group the right concordances with the appropriate pattern? How?*, should include more detailed explanations or examples to guide students in understanding the task. Similarly, the instructions in activity C, *Identify what types of words usually follow them, such as nouns, verbs, collocations, adverbs*, require further clarification. Technical terms like *nouns, verbs, collocations, and adverbs* should be explicitly explained to ensure that students understand the task requirements.

In addition, the instructions in activity C, *Do you think the words that co-occur with the lexical bundles mentioned play an important role in the sentence and indicate specific actions in engineering?*, need to be revised to make the questions more specific. I should clearly point out the types of words students should focus on and provide guidance on how to analyse them. These adjustments are necessary because the questions in their current form may be too difficult for engineering students to answer without additional support. Other minor points include addressing grammatical mistakes and improving the formatting of the materials to enhance clarity and consistency.

The comments from piloting a previous version of the materials were justified and used to improve and revise the materials. The revised materials, developed based on feedback from the piloting phase, are included in Appendix 15 of this thesis. These materials were updated to incorporate student feedback, such as clearer explanations, revised sequencing of activities, and improved task clarity. By including the revised materials in the appendix 15,

readers can evaluate the adjustments made and understand how the materials evolved through the piloting process.

The revised materials have been thoughtfully designed to scaffold students' understanding of lexical bundles while fostering critical thinking skills through exploration, comparison, and production. The first part, Presentation, introduces foundational concepts such as a corpus and concordance lines, helping students analyse patterns and contextual meanings. By explaining what a corpus and concordance are, students are exposed to linguistic tools used for analysing authentic language use, which is essential for understanding lexical bundles. The group activity requires students to observe, analyse, and infer patterns in concordance lines, encouraging analytical reasoning and pattern recognition. Additionally, identifying the types of words before and after the node word (e.g., *house*) enables students to understand the structure and function of lexical bundles in context. The collaborative learning aspect of group discussions and class-wide sharing further promotes deeper engagement with the material.

The second part, Practice, deepens students' understanding by having them compare the usage of lexical bundles in spoken and written academic contexts using different corpora, such as the British Academic Written English Corpus (BAWE) and the British Academic Spoken English Corpus (BASE). By comparing concordance lines from these corpora, students critically evaluate how lexical bundles vary in different contexts, broadening their understanding of appropriate usage. Questions like *Does the usage differ in spoken vs. written corpora?* require students to synthesize observations and formulate evidence-based conclusions, enhancing their critical thinking skills. Observing differences between formal written contexts and informal spoken contexts teaches students how lexical bundles adapt to situational demands, an important skill for effective communication. Allowing students to write their observations in Thai or English ensures inclusivity and bridges the gap between their native language and English academic language.

The third part, Production, introduces students to hands-on application by using AntConc to explore the Engineering Research Articles (ERAs) Corpus. This section equips students with the ability to independently analyse lexical bundles using corpus tools and

apply this knowledge to engineering-specific contexts. Tasks such as identifying patterns of usage and comparing meanings in general academic contexts versus engineering contexts require students to evaluate and interpret language data, promoting higher-order thinking. Additionally, analysing the lexical bundle *current* in the context of engineering research articles helps students understand its forms, functions, and meanings within their academic field. The reflection task encourages metacognitive thinking as students consider how lexical bundles in engineering contexts differ from general academic use, enabling them to internalize discipline-specific language norms.

Overall, this activity combines foundational knowledge, applied practice, and independent production to enhance students' understanding and use of lexical bundles. It fosters critical thinking by engaging students in observation, analysis, and synthesis, ensuring they not only learn about lexical bundles but also develop the skills to use them effectively in academic and professional writing. The gradual progression from understanding to practicing and producing ensures that students build confidence and competence step-by-step, aligning with the goal of empowering them to become autonomous learners and effective academic writers. While students with limited experience in corpus linguistics may initially struggle, the step-by-step guide ensures accessibility, and dividing the tasks across multiple sessions mitigates time constraints. This activity demonstrates the potential for combining lexical bundle instruction with critical thinking development in English for Specific Purposes (ESP) contexts.

4.2.4. Implementation of the workshop

In this section, there are four main components of implementing the workshop: participants and ethical approval, workshop programme, lexical bundles introduction, teacher's role and language use in the classroom.

4.2.4.1. Participants and ethical approval

The ethical forms and documents related to the project have been reviewed and approved by the FASS-LUMS Research Ethics Committee **(**FASS LUMS REC**)**, Lancaster University**.** The approved ethics documents and project proposal were handed to the Dean of the Faculty of Science and Engineering**.** Engineering students were recruited according

to the criteria specified in the proposed project. For example, students should be in the third year or fourth year of their study, their previous English grades should be from D, including D+, C, C+, B and B+, to A and they must attend all workshop sessions. All the undergraduate students were in their twenties. The students who volunteered to join the workshop signed a consent form before attending it (see participant information sheet in Appendix 5. Twenty-four participants from electrical engineering signed up, and only 13 students attended all sessions. So, the data (pre-test, post-test, students' notes, student interviews from 13 students were used in the data analysis. The students had mixed ability in English. They had passed all the compulsory English courses provided at KU, including Foundation English (I, II, III) and English for Specific Purposes (IV) (see Appendix A). Their past English grades (Foundation English III, English for Specific Purposes IV) ranged from D, through D+, C, C+, B and B+, to A. Their levels of English proficiency were elementary, pre-intermediate and intermediate. They had never been exposed to concordance lines or data-driven learning before. The participants had good computer literacy. The students' details are summarised in **Error! Reference source not found.** below.

No.			Undergraduate	Past English Grade	Age
	Initials	Gender	Year	(English III or IV)	Range
S1	K. U.	Male	4th	B+/B	21-22
S2	J. J.	Female	4th	C/B+	21-22
S3	N. Y.	Male	4th	D/B	21-22
S4	N. J.	Female	4th	B/B+	21-22
S5	P. S.	Female	4th	C/B	21-22
S6	P. P.	Male	4th	D/B+	21-22
S7	W. K.	Male	4th	B/A	21-22
S8	W. S.	Female	4th	C/B	21-22
S9	S. K.	Female	4th	C+/A	21-22
S10	S.-U. K.	Female	4th	D/C+	21-22
S11	M. U.	Male	4th	D+/C	21-22
S12	S. P.	Female	4th	D+/C+	21-22
S13	A. H.	Male	4th	D/D	21-22

Table 16 Participant details

4.2.4.2. Workshop programme and use of the materials

The workshop programme implemented was part of an English training programme provided by the Department of English at KU.CSC, Thailand. The workshop was a cooperation between the Faculty of Liberal Arts and Management Science and the Faculty of Science and Engineering between February and March 2016. The Dean of the Faculty of Science and Engineering considered there was an urgent need for engineering students to develop English in their specialist areas. The lexical bundles workshop would be a good place to start teaching writing engineering research articles. The lexical bundles used in the engineering research articles workshop covered the language usage of lexical bundles used in engineering research articles. The overall aim of the workshop was to train engineering

students in lexical bundles' forms and functions and to study lexical bundles in the engineering research articles (ERAs) corpus with the aid of AntConc concordances.

The materials used in this workshop were designed and finalized after the pre-test was administered during the first session of the workshop (Week 1). These materials were developed to specifically target the forms and functions of lexical bundles and were used as part of the teaching sessions (Weeks 2 to 4). In these sessions, students studied lexical bundles in the engineering research articles (ERAs) corpus using the AntConc concordance tool. The students were required to upload the ERAs corpus provided to desktop machines and then use the data in all the training sessions. Only the students who attended the workshop were allowed to access the computer laboratory provided by Kasetsart University for the workshop sessions. The students had a desktop computer each and were encouraged to discuss the activities with one another. The programme lasted for 10 hours in total. Week 1 was an introductory session and included taking a pre-test, while in each of weeks 2 to 4 they were trained in lexical bundles' forms and functions in conjunction with a specific corpus tool, AntConc concordances. The materials were used during this phase of the workshop and directly contributed to the students' learning experience. In week 5, the students took a post-test, and they were asked to evaluate the programme in individual interviews. The workshop programme is summarised in **Error! Reference source not found.** below.

Week	Time (hrs)	Topic	Focus	Corpus procedure/ tool/ materials
1	1.5	Pre-test (1.5 hours)		Pre-test
2	2.5	Session 1: Introduction to the concept of lexical bundles in ERAs	Individual words, e.g., the word 'house'	AntConC concordance/corpus-based activities
3	2.5	Session 2: Patterns of lexical bundles in ERAs	1. Lexical bundles' forms and functions	AntConC concordance/corpus-based activities
	2.5	Session 3: Functions of lexical bundles in ERAs	2. Noun phrases and clauses 3. Collocations of selected bundles 4. Functions of selected words e.g.,quantity (number of), cause and effect (due to)	
4	2.5	Session 4: Draw a conclusion from what they have been trained in		
		Inform students of the objectives and content of the project		
5	1.5	Post-test		Post-test
	3	Evaluate the materials and the workshop		Student individual interviews

Table 17 Workshop programme

4.2.4.3. Lexical bundles introduction: structures and functions/ AntConc concordances

This section discusses the procedures used to introduce lexical bundles and AntConc concordances in the workshop. At the very beginning of the session (see Part 1 of Session 1 in Table 18), the teaching points were the concepts of corpora and concordances. First,

the students were encouraged to think about their background knowledge and corpora as a whole class. Then, the students were introduced to loading and installing AntConc onto their desktop machines. The students were made familiar with concordances by studying the word 'current' in concordances (see materials in Part 2 in Appendix M).

Materials session	Contents	Objectives
Session 1	1. Corpora and AntConc concordances introduction	1. Corpora and concordance knowledge assessment 2. AntConc concordance tool, introduction and preparation 3. Students' attitudes towards English and corpus-based activities
Session 2	1. Lexical bundles' forms and functions 2. Noun phrases and clauses 3. Collocations of selected bundles 4. Functions of the selected words, e.g.,quantity (number of), cause and effect (due to)	1. Lexical bundles' forms and functions introduction: in order to (358), due to the (276), the number of (202), the use of (161), the effect of (139)

In session 2, the patterns and functions of lexical bundles were introduced. The introduction of lexical bundles' forms followed the steps of a well-established structure for teaching in the classroom, present (P), practice (P) and production (P), known as PPP (Ur, 1996).

In the presentation stage, the concepts of the lexical forms and functions of lexical bundles were introduced by asking students to underline lexical bundles and identify their

position in engineering texts within research articles: at the beginning, in the middle, or at the end. For example, students were guided to observe where the noun phrase *the number of* appears within the concordance lines. This activity helps students recognise how the position of lexical bundles can vary depending on the sentence structure and function. Understanding these positions is important because it allows students to see how lexical bundles are used in context, which can improve their ability to use them accurately in their own academic writing.

It is important to teach students how to identify the positions of lexical bundles in sentences because it helps them understand how these bundles function within the overall sentence structure. The lexical bundles are considered the building blocks of academic writing, as they help students understand how phrasal units, such as *the number of*, connect to other parts of a sentence. For example, in the sentence *I have installed solar cells, the number of three kilowatts*, the phrase *the number of* is categorised as a noun phrase with *of*, and it connects to *three kilowatts*. This demonstrates the pattern *noun of noun*, where *three kilowatts* is the noun connected to the bundle. In this case, *the number of* is positioned at the end of the sentence. However, students can also be shown how this same bundle can appear in other positions. For instance, *The number of three kilowatts was installed* places the bundle at the beginning of the sentence, while in another case, it could be positioned in the middle of a sentence. Teaching these variations is important because it allows students to notice the positional flexibility of lexical bundles while working with concordances. Understanding these patterns not only enhances students' awareness of how lexical bundles function but also helps them deduce the meaning of words in context. This is particularly useful for engineering students, as recognising these structures can assist them in accurately interpreting and constructing sentences in their academic and technical writing.

The students then moved on to exercises in order to practise the forms of lexical bundles**.**

In term of the functions of lexical bundles, the students were asked to study AntConc concordances and then interpret the meaning of each bundle**.** The students were given examples of how to read, form hypotheses and interpret concordances **(**see Appendix

M**).** Then, the students carried out the same steps with other bundles**.** In the production stage, the students were asked to produce a paragraph about their research or a project they were working on by using the lexical bundles provided **(**see **"**Over to you" section in the materials, Appendix M**).**

The format of the class was divided into three stages**:** the facilitator informed the students about the focus of the lesson, the facilitator and the students did the activities as a whole class, the students worked at their own pace or they could discuss with their f concordances to teach lexical bundles is based on the forms and meanings of lexical bundles used in ERAs. The materials produced were employed in the teaching.

4.2.4.4. Teaching of lexical bundles

Based on the above discussion, I organised the teaching into two types of activities: introducing the corpus and using AntConc concordances, and teaching the forms and functions of lexical bundles. It should be noted that the facilitator (me) and the students worked together in class. The activities provided required the students to contribute by sharing answers with their friends in the classroom (see Appendix 8).

The first activity was to introduce the students to corpora. I presented the words 'house and housing',which are in concordances, and asked the students 'What are they?' and 'What do these concordances represent?' Then, the students found the answer that the words 'house and housing' are used in concordances. Then, I linked the meanings of concordances to the concept of 'corpora' and the benefits that could be achieved in the workshop. Since AntConc is used as the primary tool for learning the bundles, the students and I worked together to download AntConc [22] and set it up before working on lexical bundles activities.

The second kind of activity focuses on teaching lexical bundles based on their structure and usage. The techniques employed in the classroom include interpreting lexical bundles from concordances, analysing their meanings in context, and identifying how words within

[22] Students worked on their own desktop machine in the computer laboratory so that they would fully benefit from using a desktop machine to work on their activities.

a span of 1to 5 words around the lexical bundles contribute to their meaning. In this case, learning involves understanding the immediate linguistic environment of the bundle, recognising patterns in the surrounding words, and how these patterns influence the lexical bundle's function and meaning within a sentence. For example, students analyse whether the words around a bundle, such as *the number of*, include nouns or other elements that provide contextual clarity, helping them better understand and apply the bundle in their own writing.

The students were explicitly encouraged to generalise the meanings of lexical bundles from concordances. When interpreting lexical bundles, the students were explicitly guided to generalise the meanings of lexical bundles by analysing their use in concordances, identifying patterns, and observing how the bundles function in different contexts. This process involved analysing multiple occurrences of a lexical bundle in different contexts presented through concordance lines. For example, when working with the bundle *in the case of*, students examined its usage in several concordance lines to identify its common functions, such as introducing an example or specifying a scenario.

By observing patterns in how the bundle was used across various sentences, students could abstract a general understanding of its pragmatic function and syntactic structure. This activity required students to focus on the recurring elements in the surrounding linguistic context, such as the types of nouns or verbs associated with the bundle, as well as its position within the sentence. Generalising the meaning of a bundle helps students move beyond rote memorization, enabling them to apply their understanding flexibly to new contexts. For instance, after analysing concordances, students might recognise that *in the case of* is versatile and can be adapted to discussions of hypothetical, real, or comparative scenarios. This skill is crucial for developing a deeper comprehension of lexical bundles and for fostering the ability to use them accurately in academic and professional writing.

But it was found that they could not interpret the meanings from concordances. Therefore, when they worked with each lexical bundle they worked with five concordances each time until they could find the meanings of lexical bundles. Also, they had to notice

the structure of lexical bundles. In terms of the words that surround lexical bundles, the students were asked to note the words that normally occur with lexical bundles (see Appendix 3).

The students then proceeded to exercises designed to practise the forms of lexical bundles. In terms of the functions of lexical bundles, the students analysed AntConc concordances and interpreted the meaning of each bundle. They were provided with examples of how to read concordances, form hypotheses, and interpret the patterns. Following this, the students applied the same steps to other bundles. They were provided with examples of how to read concordances, form hypotheses, and interpret patterns. For instance, when analysing the concordance *the number of*, students were shown concordance lines such as:

- *The number of experiments conducted was limited due to time constraints.*
- *We increased the number of participants to ensure better accuracy.*
- *The number of variables in the study significantly impacted the results.*

Through these examples, students were guided to observe the context in which *the number of* appears, identify its grammatical structure as a noun phrase with *of*, and examine the types of words that typically follow it (e.g., *experiments*, *participants*, *variables*). They were then encouraged to hypothesise that *the number of* often introduces a quantity or count related to specific nouns, serving a descriptive function in academic writing. This process helped students interpret the concordance patterns and apply similar analyses to other lexical bundles.

During the production stage, the students were tasked with writing a paragraph about their research or a project they were working on, incorporating the lexical bundles provided (see the *Over to you* section in the materials, Appendix 13). The *Over to You* section includes activities designed to improve students' understanding of lexical bundles and their applications in engineering texts. In *Activity 2*, students underline the subjects and objects in sentences, noticing patterns such as subjects being single nouns or including modifiers and objects often being identifiable through specific linguistic structures. For example, in the sentence *The permeability is increased from approximately 3.6 x 10^-10*

cm/s to around 13 x 10^-10 cm/s when VPA content is increased, students observe the subject *permeability* and the object *VPA content*. In *Activity B*, students examine how lexical bundles like *the effect of* and *the number of* function in research articles, grouping concordances into patterns based on their position and usage. For instance, in *The effect of pile spacing on thermally induced mechanical behaviour*, students note how *the effect of* introduces a prepositional phrase. This helps students identify bundles in different sections of research articles and understand their contextual roles. In *Activity C*, students focus on co-occurring words with bundles like *the use of* and *the number of*, categorising them into nouns, verbs, or adverbs. For example, in *The number of buffers improves the power efficiency*, students classify *buffers* as a noun and discuss its role in quantifying technical elements. These activities culminate in a table where students summarise their observations. Additionally, a self-study section provides foundational knowledge about phrases and clauses, teaching students to distinguish between structures like *the effect of percentage of VPA* (a phrase) and *In order to better capture the effect of the segment* (a clause). Together, these activities enhance students' ability to recognise and interpret lexical bundles, analyse their functions, and apply them in their own academic writing. By engaging with authentic engineering texts and concordance data, students develop practical analytical skills essential for effective technical communication.

The format of the class was divided into three stages: first, the facilitator introduced the focus of the lesson; next, the facilitator and students completed the activities together as a whole class; finally, the students worked independently or discussed with their peers at their own pace. The use of concordances to teach lexical bundles focused on the forms and meanings of lexical bundles found in Engineering Research Articles (ERAs). The materials developed for this purpose were employed throughout the teaching process.

4.2.4.5. Teacher's role and language use in the classroom

The teacher's role was to be a facilitator who monitored and worked with the students. The facilitator guided the students and provided help if the students needed it. The main languages used in the classroom were Thai and English. The facilitator only used Thai to explain the concepts of lexical bundles that seemed to be problematic to the students. The

use of students' first language (L1) in EFL instruction has been widely debated in the field of language teaching. Proponents of L1 use argue that it can serve as a cognitive scaffold, facilitating comprehension of complex L2 concepts, reducing learner anxiety, and enabling efficient classroom management (Butzkamm & Caldwell, 2009; Hall & Cook, 2013). For instance, Butzkamm and Caldwell (2009) advocate for *enlightened monolingualism,* emphasizing that strategic L1 use helps bridge the gap between learners' existing knowledge and new linguistic inputs. Hall and Cook (2013) also highlight the practical benefits of L1 in explaining grammar and managing classrooms, aligning with the realities faced by many language teachers. However, critics of L1 use caution that it may hinder immersion in the target language (L2) and reduce the amount of L2 exposure learners receive (Ellis, 2005). This perspective aligns with the monolingual principle, which emphasizes maximizing L2 use for more effective language acquisition. Recent empirical studies, such as Yüzlü and Atay (2020), offer nuanced insights, suggesting that L1 use can promote L2 production, particularly for lower-proficiency learners, by providing a cognitive bridge to new linguistic concepts. These diverse perspectives suggest that the effectiveness of L1 use in instruction depends on factors such as learners' proficiency, instructional goals, and the specific context of the learning environment.

In the classroom, the facilitator monitored the students' use of the Internet while doing the activities. The students had to follow the workshop rules. The rules were that they could access a computer in the computer laboratory for the purpose of accessing the ERAs corpus with AntConc (see Anthony, 2011), but they were not allowed to surf the Internet. The decision to ban internet use by students during certain activities was made to ensure that their understanding and application of lexical bundles were based on the materials and guidance provided in the study rather than external resources. This restriction was relevant for several reasons. First, it ensured that students focused on the specific learning objectives of the study without being distracted by unrelated information or overwhelmed by the vast array of resources available online. Second, it helped create a controlled environment where the effectiveness of the teaching materials and strategies could be accurately assessed, free from external influences. Third, banning internet use

encouraged students to rely on their analytical and problem-solving skills, such as deducing the meaning and function of lexical bundles from provided concordances, rather than resorting to quick online searches. This approach was intended to foster deeper learning and independent thinking, which are critical for academic writing. Finally, the restriction reduced the risk of students encountering inaccurate or irrelevant information that could potentially mislead them. While internet resources are valuable, their selective exclusion in this context served to maintain the integrity and focus of the learning process.

4.2.4.6. Analysis of students' use of lexical bundles

The students' writing activities 2 in the pre-test and post-test were analysed regarding the use of lexical bundles in writing production**.** In particular, the analysis focused on grammatical errors related to the utilisation of lexical bundles and grammatical errors are unrelated to lexical bundle**.** The analysis focused on identifying and categorising grammatical errors made by students in their writing, specifically distinguishing between errors related to the use of lexical bundles and those unrelated to lexical bundles. Grammatical errors related to lexical bundles included issues such as incorrect word order, missing components (e.g., omitting necessary prepositions in bundles like *in the case of*), or misuse of a bundle in a context where it did not fit syntactically or pragmatically. These errors reflected a lack of understanding of how lexical bundles function within sentences.

In contrast, grammatical errors unrelated to lexical bundles encompassed issues like subject-verb agreement, verb tense errors, or inappropriate use of articles, which were more general indicators of students' overall grammatical competence. The purpose of analysing both types of errors was to gain a comprehensive understanding of the challenges students faced in their academic writing. By distinguishing between these categories, the study aimed to assess whether targeted instruction on lexical bundles improved students' ability to use them accurately, while also identifying areas where additional grammar support might be needed.

4.2.5. Data collection

4.2.5.1. Research instruments

There were five research instruments used in the data collection**:** pre**-**test and post**-**test, student interviews, classroom recordings, students' notes and field notes. The purpose was to see whether the lexical bundles workshop helped to improve the students' research article writing. An inter-rater reliability check was carried out to check the validity of the 13 students' writing results in writing part 2 and the categories of grammatical errors that occurred in the students' writing.

4.2.5.1.1. Pre-test and post-test

The pre-test and post-test were divided into two parts: understanding the meaning and structure of LBs and paragraph production. The test lasted for one and a half hours (1 h 30 m).

Part 1 consisted of multiple-choice questions (10 items, max. score 10) based on the use of lexical bundles in sentences. The second part asked students to write a paragraph by using lexical bundles provided (1 item, max. score 10). These types of exercises were adapted from those developed for teaching collocations (Conzett, 2000; Hill, Lewis & Lewis, 2000; Lewis, 2000).

The writing section in part 2 involved writing a paragraph (1 item, max. score 10). Topics included *the new technology used for saving power or with environmental benefits*, *an engineering laboratory report, the weak points of having a too advanced technology, or the future of technology adapted for smart homes.* A lexical bundles list was provided for students to choose from when writing. Based on the lexical bundles used in producing the paragraph, students could use five lexical bundles as a maximum. This limit was set to encourage students to focus on the quality rather than the quantity of their usage. By restricting the number of bundles, students were required to select the most contextually appropriate and functionally effective bundles for their paragraphs. This approach aimed to avoid overuse or redundancy, which could lead to unnatural or incoherent writing. Additionally, limiting the number of bundles helped students pay closer attention to their grammatical accuracy, placement, and pragmatic relevance within the paragraph, fostering

a deeper understanding of how lexical bundles contribute to academic writing. Practically, this restriction also allowed for more manageable evaluation by instructors, as each bundle's usage could be thoroughly assessed for correctness and appropriateness. Moreover, a maximum of five bundles aligns with findings from corpus studies, which indicate that high-quality academic writing tends to employ a moderate and strategic use of lexical bundles, rather than an excessive number. The correct use of lexical bundles in writing was evaluated, with each lexical bundle earning a maximum of 2 marks based on how well it matched the syntax of a complex sentence. The students were asked to write about different topics related to science and engineering.

A t-test (one-tailed) was used to measure the students' improvement in using lexical bundles in engineering research articles writing and determine whether the training in the use of lexical bundles in writing engineering research articles was significant. A paired t-test was appropriate, each person was paired with him- or herself on the two tests.

4.2.5.1.2. Student interviews

The interview questions were based on the teaching of lexical bundles, the workshop and students' attitudes towards the workshop (see Appendix K). The students were interviewed individually at the end of the workshop. Seven questions were asked as follows:

Question 1: What do you think about the training sessions? (Did you like them? How useful are they to you?)

Question 2: What aspects of the training sessions did you like most (e.g.,learning materials, content etc.)? Why?

Question 3: How useful are the materials?

Question 4: How useful is the teaching of lexical bundles (fixed sequences of words)?

Question 5: To what extent do you think the ERA sessions help to improve your English skills? And in what aspects (e.g.,word chunks, grammar, syntax, semantics)?

Question 6: Overall, do you think the training sessions help to improve your ability to write engineering articles?

Question 7: Do you have any comments on the training sessions?

The feedback questions used in the study primarily focused on students' positive experiences with the teaching materials and methods. It is acknowledged that the questions were somewhat similar in nature and did not explicitly invite students to discuss aspects they disliked or found challenging. This may have limited the scope of feedback to predominantly positive aspects. Including more diverse questions, such as asking about difficulties or areas for improvement, could provide a more balanced evaluation of the teaching methods. This limitation is addressed in the conclusion section as a recommendation for future studies.

4.2.5.1.3. Classroom recordings

All teaching sessions in the workshop were recorded.

4.2.5.1.4. Field notes

The researcher made detailed notes after the sessions. The notes consisted of a summary of what happened in the classroom, and what the teacher thought about the class, such as students' progress and students' reactions to the class.

4.2.5.1.5. Students' notes

The students' notes were collected in order to consider the students' attitudes towards the workshop, and how well the students understood lexical bundles.

4.2.5.2. Inter-rater reliability

An evaluation of inter-rater reliability was carried out to enhance the validity and reliability when marking students' writing (Mackey & Gass, 2005). There were two main parts in the pre-test and post-test: multiple choice questions and writing a paragraph. In part 1, the activity concerned multiple choices for choosing the right lexical bundles to fill in the gaps. Students were given a score of zero if they chose a wrong lexical bundle, and a score of one (1 score) if they chose the right lexical bundle to fill in a gap. Checking inter-rater reliability was not necessary in activity 1 because the researcher checked the correct

answers in the ERAs corpus. Instead, inter-rater reliability checking was applied to the students' writing part 2 from the pre-test and post-test. The three raters were a doctoral researcher and doctoral students in Applied Linguistics. They have experience in teaching English for EFL/ESL students. In part 2, the inter-rater reliability consisted of two stages.

4.2.5.2.1. Stage 1: Judgement on the students' use of lexical bundles

The first stage was a judgement on the students' use of lexical bundles in the pre-test and post-test (§ 4.2.4.2). There were two main stages in checking reliability in the students' writing tests. First, the criteria set were based on the students' production of lexical bundles. In other words, the lexical bundles and the words that followed them should agree regarding grammatical properties. In this stage, the three raters were trained in how to mark students' writing according to the students' use of lexical bundles mentioned to able to reach an agreement. Each rater scored all 13 students' written work individually. Second, all the marking from the three raters was coded. In order to do the coding, two coding systems were used for correct and incorrect answers in the students' writing tests. The marking results were divided into two sections, from the pre-test and the post-test. The researcher coded 1 if all three raters agreed that the answers in the students' writing were correct and incorrect, as shown in Table 18 and Table 19.. A code 0 was given if one or two of the raters did not give the same answers. To be able to arrive at the same research findings, the threshold level for agreement among the three raters was set at 90 per cent. **Error! Reference source not found.** and **Error! Reference source not found.**, below, show the number of correct and incorrect uses of lexical bundles judged by the three raters in the pre-test and post-test.

After the raters had marked the students' writing, the three raters compared and discussed their judgements of the scores. If there were points on which the raters did not agree, negotiations took place until agreement was reached.

Student No.	Pre2 scores	Correct	Incorrect	Rater 1 correct	Rater 2 correct	Rater 3 correct	Coding	Rater 1 incorrect	Rater 2 incorrect	Rater 3 incorrect	Coding
S1	0	0	2	0	0	0	1	2	2	2	1
S2	0	0	1	0	0	0	1	1	1	1	1
S3	0	0	0	0	0	0	1	0	0	0	1
S4	0	0	1	0	0	0	1	1	1	1	1
S5	0	0	1	0	0	0	1	1	1	1	1
S6	0	0	2	0	0	1	0	2	2	1	0
S7	4	2	1	2	2	2	1	1	1	1	1
S8	2	1	0	1	1	1	1	0	0	0	1
S9	0	0	2	0	0	0	1	2	2	2	1
S10	2	1	0	1	1	1	1	0	0	0	1
S11	0	0	0	0	0	0	1	0	0	0	1
S12	0	0	2	0	0	0	1	2	2	2	1
S13	0	0	0	0	0	0	1	0	0	0	1
	Percentage of agreement						92.31 %				92.31 %

Table 18 Coding and percentage agreement in the writing pre-test part 2

The negotiations of the three raters about judging the writing of student S6 is shown below. For example, the judgements of the correct use of the lexical bundle ‘the number of’ from the student did not match. One of the raters marked the use of the bundle ‘the number of’ as correct, the other two raters did not agree. Three of the raters had to check the student’s usage of this bundle against the usage of the bundle ‘the number of’ in the concordances. Based on the concordances in the ERAs corpus, we agreed that the use of the bundle ‘the number of’ in the student’s writing was incorrect.

> … 'But the number of many students, makes me match with someone students.' 'I want the formation of my family.'… (Excerpt from S6’s writing pre-test)

Regarding meaning, the use of the words ‘many students’ after the bundle ‘the number of’ was understandable. The quantifier ‘many’ is grammatically correct to use with the plural noun ‘students’ (Quirk et al., 1985).

No.	Post 2 scores	Correct	Incorrect	Rater 1 correct	Rater 2 correct	Rater 3 correct	Coding	Rater 1 incorrect	Rater 2 incorrect	Rater 3 incorrect	Coding
S1	2	1	0	1	1	1	1	0	0	0	1
S2	6	3	2	3	3	3	1	2	2	2	1
S3	4	2	0	2	2	2	1	0	0	0	1
S4	4	2	0	2	2	2	1	0	0	0	1
S5	6	3	1	3	3	3	1	1	1	1	1
S6	4	2	1	2	1	1	0	1	1	1	1
S7	6	3	1	3	3	3	1	1	1	1	1
S8	4	2	0	2	2	2	1	0	0	0	1
S9	10	5	0	5	5	5	1	0	0	0	1
S10	6	3	0	3	3	3	1	0	0	0	1
S11	0	0	2	0	0	0	1	2	2	2	1
S12	8	4	0	4	4	4	1	0	0	0	1
S13	2	1	1	1	1	1	1	1	1	1	1
	Percentage of agreement						92.31 %				100%

Table 19 Coding and percentage agreement in the writing post-test part 2

However, the raters agreed that the utilisation of the word 'many' after the lexical bundle 'the number of' was incorrect. So, student (S6) did not score one (1) for this item. The numbers of correct and incorrect uses of lexical bundles were calculated as percentages, as shown in Table 18 and Table 19 for percentage agreement (Mackey and Gass, 2005: 243). The results of coding were calculated as percentages of agreement. The inter-rater reliability of the students' correct and incorrect use of lexical bundles in the pre-test was 92.31% (see **Error! Reference source not found.**). The inter-rater reliability of the students' correct and incorrect use of lexical bundles in the post-test was 92.31% and 100% (see **Error! Reference source not found.**).

4.2.5.2.2. Stage 2: Judgement of the students' grammatical errors related to or not related to the use of lexical bundles

The sentences produced by students in the pre-test (31 in total) and post-test (50 in total) were divided into two groups.The first one consisted of sentences related to lexical bundles, while the second one did not include lexical bundles in writing. The three raters identified the grammatical errors and meanings in the students' sentence production separately. There were lists of grammatical errors provided as guidelines, and the raters were able to add other grammatical errors apart from those on the lists. The three raters considered the meanings of sentences produced concerning how the students conveyed comprehensible or incomprehensible ideas based on their use of lexical bundles (79 in total).

Regarding agreement, all sentences analysed by the three raters were identified and compared until a consensus was reached. The researcher coded 1 if all the three raters assigned the same results. A coding of 0 was given if one or two of the three raters did not give the same answers. The percentage of agreement sought was 95 per cent among the three raters (see Appendix I).

4.3. Findings and discussion

This section aims to address the effectiveness of the workshop, including the teaching of lexical bundles and creating materials for engineering students in conjunction with the use of data-driven learning in the classroom (to answer research question 3). There are four main sections presented. Section 4.3 reports the findings from the pre-test and post-test scores. In particular, this study attempts to categorise and identify grammatical errors related to the use of lexical bundles in the pre-test and post-test (§ 4.3.2.1). The results in Section (4.3.2.2) also briefly mention grammatical errors not related to the use of lexical bundles. The effectiveness of the study is reflected in the extension of the students' vocabulary knowledge after learning lexical bundles in the workshop (§ 4.3.3). The materials development process and the application of data-driven learning (DDL) in the study are also discussed (§ 4.3.4.4). The results from this section have methodological and

pedagogical implications and point the way to further research in the area **(**see also sections 5**.**3**.**1 and 5**.**3**.**2 in Chapter 5**).** A summary of the chapter is in Section 4**.**4**.**

4.3.1. Findings from the pre-test and post-test

The table provides details about 13 fourth-year undergraduate students, aged 21–22, enrolled in an English course. The group consists of both male and female students, with a majority being female. Their past English grades range from D to A, reflecting varying levels of proficiency. Pre-test and post-test scores were recorded to assess improvement in their English skills. Most students showed significant progress, with score increases ranging from +1 to +11. For instance, S12 (female) demonstrated the highest improvement, scoring 4 in the pre-test and 15 in the post-test (+11), while S6 (male) showed the lowest increase of +1.

In this section, the results from both parts of the pre-test and post-test are presented. Each part had a maximum score of 10. As shown in Table 20, all 13 students demonstrated significant improvement, with their total scores in the post-test at least doubling compared to their pre-test scores.

No.	Initials	Gender	Undergraduate	Past English Grade	Age	Test Results		
			Year	(English III or IV)	Range	Pre-Test	Post-Test	Increase/Decrease (+/-)
S1	K. U.	Male	4th	B+/B	21-22	2	5	+2
S2	J. J.	Female	4th	C/B+	21-22	3	9	+6
S3	N. Y.	Male	4th	D/B	21-22	2	6	+4
S4	N. J.	Female	4th	B/B+	21-22	2	8	+6
S5	P. S.	Female	4th	C/B	21-22	1	9	+8
S6	P. P.	Male	4th	D/B+	21-22	5	6	+1
S7	W. K.	Male	4th	B/A	21-22	6	9	+3
S8	W. S.	Female	4th	C/B	21-22	5	10	+5
S9	S. K.	Female	4th	C+/A	21-22	2	12	+10
S10	S.-U. K.	Female	4th	D / C+	21-22	4	9	+5
S11	M. U.	Male	4th	D+/C	21-22	2	5	+3
S12	S. P.	Female	4th	D+/C+	21-22	4	15	+11
S13	A. H.	Male	4th	D/D	21-22	2	4	+2

Table 20 Descriptive data about the 13 participants

	Pre-test	Post-test
N	13	13
Valid		
Missing		
Mean	1.23	1.69
Std. Deviation	0.52	0.87
Minimum	1.00	1.00
Maximum	5.00	7.00

Table 21 Pre-test and post-test from part 1

In part 1, there were ten multiple choice questions and the maximum score was ten. The purpose was to focus on the correct use of lexical bundles in the sentences. The sentences were taken from concordances in the ERAs corpus. **Error! Reference source not found.** show that eight students increased their scores between the pre-test and the post-test: S1, S2, S4, S5, S8, S10, S11 and S12. Only three students had the same scores: S3, S9 and S13. And S6 and S7 had lower scores in the post-test. The difference between the pre-test and post-test in part 1 was not statistically significant (t-test = .082, $p < .050$). Looking at the results in more detail, these results suggest that students (S1, S2, S4, S5, S8, S10, S11 and S12, who have quite good grades in previous English courses, could potentially have good vocabulary knowledge or participants' descriptive data).

In contrast, the students who got lower scores in the post-test were more likely to lack vocabulary knowledge in the field and were less proficient in English in general. It can be inferred that the students' variable factor, i.e., English background, might have had an influence on doing the activities provided. It can be summarised from the results of the pre-test in part 1 that there was no major improvement in the students' knowledge of the correct use of bundles in engineering research articles after instruction. It is revealed that the students are familiar with the target bundles and can select the correct answers in part 1. Also, the students can guess from the context the target lexical bundles that are used in part 1. It is interesting but unsurprising that the students' scores in activity 1 do not change much from the pre-test to the post-test since they seem to be proficient in receptive skills rather than productive skills. For the writing skills in part 2, this involved lexical knowledge and grammatical knowledge to produce their writing.

The activity in part 2 involved writing a short paragraph (max. score of 10). Students chose lexical bundles from the list provided to write about given topics, i.e., *the new technology used for saving power or with environmental benefits, an engineering laboratory report, the weak points of having a too advanced technology* or *the future of technology adapted into smart homes*. As shown in Table 22, there was a statistically significant difference in this part (t-test = .000, $p < .001$). The findings showed that most students improved their scores in writing part 2, with the exception of one student, S11,

who achieved the same score in both the pre-test and post-test. The students who scored zero in the pre-test either did not produce any written content or created lexical bundles with incorrect forms and meanings.

Three students **(**S7, S8 and S10**)** gained scores in writing part 2 in the pre-test**.** The rest of the students were marked 'zero' in the writing pre-test part 2**.** The students more than doubled their pre-test scores in the post-test**.** For example, S9 scored zero in the pre-test but increased this to 10 in the post-test**.** Only one student, S11, out of 13 did not achieve any score in this activity**.**

	Pre-test	**Post-test**
N		
Valid	13	13
Missing		
Mean	0.62	4.77
Std. Deviation	1.26	2.65
Minimum	0.00	0.00
Maximum	4.00	10.00

Table 22 Pre-test and post-test from part 2

	Paired differences			**t**	**df.**	**Sig. (2-tailed)**
	Mean Difference	Standard Deviation	N			
Pre-test - Post-test	5.000	3.18	13	5.65	12	0.000***
Pre-test - Post-test (part 1)	.46	.87	13	1.89	12	10.082
Pre-test-Post-test (part2)	4.15	2.76	13	5.41	12	0.000***

Table 23: Paired samples t-test between pre-test and post-test

$p < 0.001$ (one-tailed)

*** indicates statistical significance p <0 .001

From Table 23, the engineering students' post-test t-test scores (part 2) were higher than their pre-test t-test scores, which means there was a positive effect for learning lexical bundles in the writing of engineering research articles.

It is suggested that, to some extent, the students were able to recognise the lexical bundles used in engineering research articles and to use lexical bundles in the writing activities provided. To some extent, the students were not able to achieve highly in the writing activity. It is possible that variable factors, such as students' proficiency level or exposure to L2, influenced the outcomes (see Paquot, 2010). However, the students' variable features are not the focus of this study, but the students' details can be referred to in this chapter. The extent to which the students could achieve or not achieve the writing activity was discussed regarding grammatical errors related to the use of lexical bundles and grammatical errors not directly linked to the use of lexical bundles (§ 4.3.2.1 and 4.3.2.2).

However, the study acknowledges that the small number of students, the variation in their ability levels, and the limited number of lexical bundles taught within a short teaching period present constraints to the generalisability of the findings. While the results indicate improvement in students' use of lexical bundles, further longitudinal studies are necessary to explore whether these improvements are retained in students' future writing.

4.3.2. Grammatical errors found in students' writing

From the results of the students' writing activity part 2 in the pre-test and post-test, it is revealed that the students could use more lexical bundles correctly and convey their ideas more clearly in the post-test than in the pre-test (§ 4.3.2.1). Alongside their progress in the use of lexical bundles, there were some grammatical errors found in the students' sentence production. This section looks at grammatical errors related to lexical bundles and grammatical errors not linked to lexical bundles in the students' writing.

As shown in Figure 5, the focus of the analysis involved grammar and meanings. On the left side of the chart, sentences were investigated to determine whether they were contextually meaningful. If the sentences were meaningful, an investigation of grammar was carried out. If grammar was correctly used in the sentences, no further investigation was done. Similarly, the sentences were explored concerning grammar, as can be seen on the right side of the chart below. Sentences that were correctly produced were considered further regarding meanings. If sentences were meaningful, there was no need to conduct any further investigation.

To sum up, sentences that were grammatically correct and meaningful were not investigated further. Instead, they were treated as correct sentences. If sentences did not convey meanings, and grammar was not used correctly, they were categorised into two groups: grammatical errors related to the lexical bundles and grammatical errors not related to lexical bundles, as illustrated in Figure 5. The analyses of the two groups of grammatical errors are described and explained in § 4.3.2.1 and 4.3.2.2.

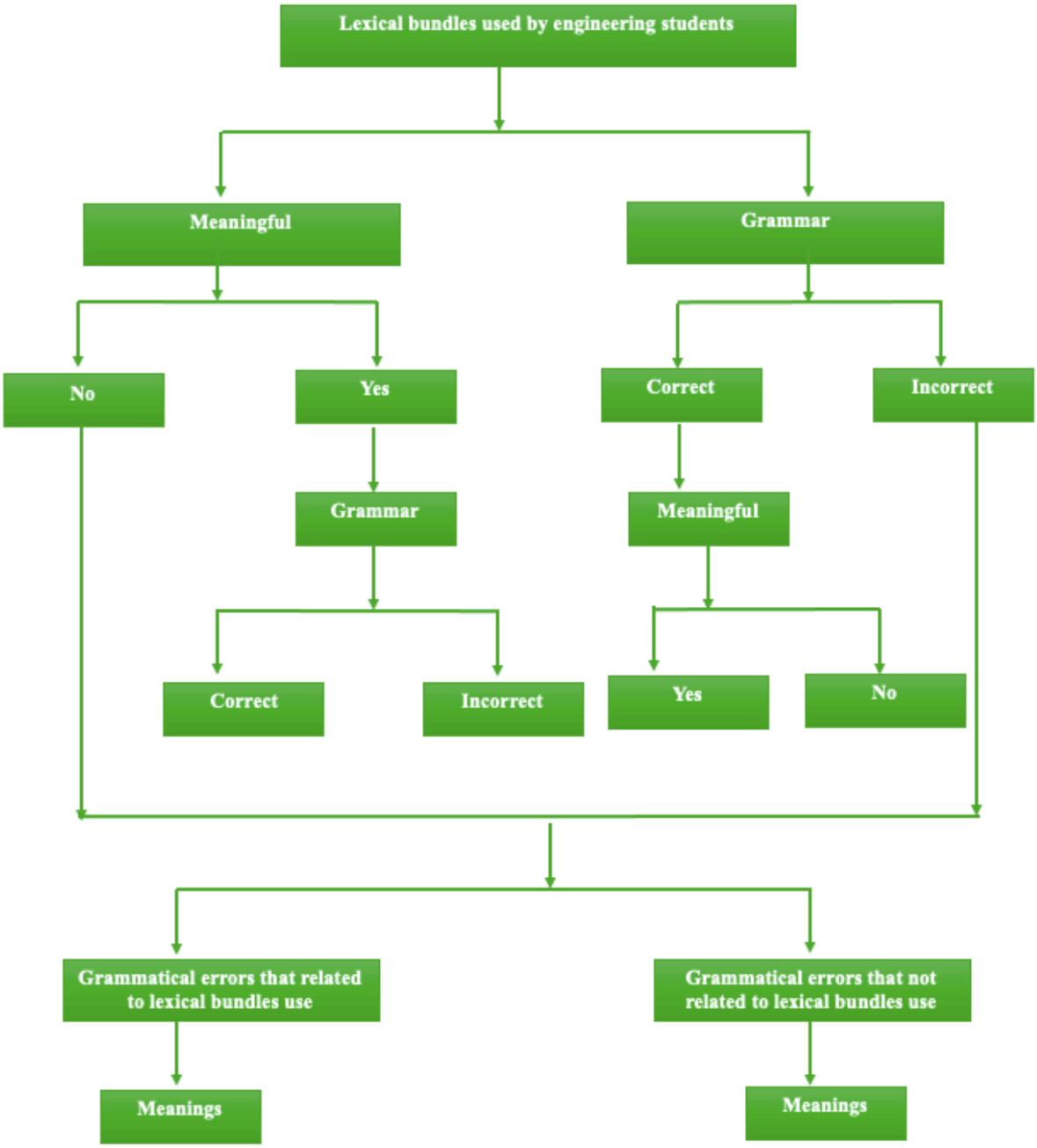


Figure 5 Analysis of lexical bundles in students' sentence production

4.3.2.1. Grammatical errors related to the lexical bundles

This section presents and highlights grammatical errors related to the use of lexical bundles in students' sentence production based on the judging of three raters. Table 25 presents a

categorisation of common grammatical and lexical errors identified in students' writing tasks, based on their performance during a specific learning activity. These examples illustrate various error types, such as incorrect lexical bundle choices, inappropriate word choices, misuse of word classes, and issues with word forms. Additionally, structural problems like run-on sentences, word omissions, fragments, and subject-verb agreement errors are highlighted. The examples, extracted directly from the students' work, provide insight into typical challenges faced when constructing written responses. This analysis aims to highlight areas requiring targeted instruction to improve their academic and professional writing skills.

Error Type	Definition	Examples from the students' writing.
Lexical Bundle Choices	Incorrect use of common word combinations or phrases that frequently occur together	The use of solar energy Instead of using the coal can be the amount of pollutants because many aspects such as the air landscapes.
Word Choices	Selection of inappropriate or incorrect words that don't convey the intended meaning	My project due to the environment of Thailand.
Word Classes	Using a word as the wrong part of speech (noun, verb, adjective, etc.)	I'm study electrical engineering due to the liked it and the effect of do it well but I'm not good English
Word Forms	Incorrect morphological form of a word (tense, plurality, etc.)	In technology design must have the effect of produce and protected system to best, future the number of many people everyone want have need in technology on intern swell up.
Run-on Sentences	Two or more independent clauses joined incorrectly without proper punctuation or conjunctions	Solar cells for study the development of solar frame but can see how to keep underneath and the formation of solar cells that as a result how, Impressions from the number of server results

Word Omissions	Missing necessary words in a sentence	The new of my produce is Energy saving lamp in this study can reduce Global warning as well as can reduce electricity charge.
Fragments	Incomplete sentences missing essential elements (subject or predicate	The use of electricity in school building.
Subject-verb Agreement	Mismatch between the subject and verb form	In order to analyse the effect of the sunlight, I used in the production of electricity in the home, it help save on energy costs.

Table 24 Categorisation of common grammatical and lexical errors identified in students' writing tasks

The table categorises and describes the common error types identified in the writing of lexical bundles by engineering students during their training. These errors were observed across multiple writing tasks and reflect challenges faced by students in applying academic writing conventions. Each error type is defined, and examples of actual student errors are provided to illustrate the specific nature of the mistakes. The aim of the table is to offer a clearer understanding of the types of linguistic issues encountered, which can inform targeted instructional interventions and materials development to improve students' proficiency in using lexical bundles effectively in academic writing. To provide more detail about the different error types identified in this study, the following descriptions have been tailored to reflect the common mistakes made by engineering students in the writing of lexical bundles.

1. **Lexical Bundle Choices**: Errors in selecting or combining words that frequently occur together in academic writing. For example, writing *"The use of solar energy instead of using the coal can be the amount of pollutants because many aspects such as the air landscapes"* instead of a clearer and grammatically correct structure like *"The use of solar energy, rather than coal, reduces pollutants and improves air quality."*

2. **Word Choices**: The use of inappropriate or vague words that fail to convey the intended meaning. For instance, writing *"My project due to the environment of Thailand"* instead of *"My project focuses on the environmental challenges in Thailand."*

3. **Word Classes**: Errors in using the wrong part of speech (e.g., noun, verb, adjective). An example is *"I'm study electrical engineering due to the liked it and the effect of do it well but I'm not good English"* instead of *"I study electrical engineering because I like it and want to improve my skills in the field."*

4. **Word Forms**: Mistakes in using the correct morphological form of a word, such as tense, plurality, or derivation. For example, *"In technology design must have the effect of produce and protected system to best, future the number of many people everyone want have need in technology on intern swell up"* instead of *"In technology design, it is important to produce and protect systems effectively to meet the needs of a growing population in the future."*

5. **Run-on Sentences**: Incorrectly joining two or more independent clauses without proper punctuation or conjunctions. For example, *"Solar cells for study the development of solar frame but can see how to keep underneath and the formation of solar cells that as a result how, Impressions from the number of server results"* instead of *"Solar cells are used to study the development of solar frameworks, focusing on how to optimise their structure and efficiency."*

6. **Word Omissions**: Missing necessary words in a sentence, leading to unclear meaning. For instance, *"The new of my produce is Energy saving lamp in this study can reduce Global warning as well as can reduce electricity charge"* instead of *"My new product is an energy-saving lamp, which can reduce global warming and lower electricity costs."*

7. **Fragments**: Incomplete sentences that lack essential components, such as a subject or predicate. For example, *"The use of electricity in school building"* instead of *"The use of electricity in school buildings is analysed in this study."*

8. **Subject-Verb Agreement**: Errors in matching the subject and verb form. For example, *"In order to analyse the effect of the sunlight, I used in the production of*

electricity in the home, it help save on energy costs" instead of *"To analyse the effect of sunlight, I used it in the production of electricity in homes, which helps save energy costs."*

Generally speaking, the results revealed that there were seven grammatical error types (in percentages) that appeared in the pre-test: lexical bundle choices (58%), word choices (58%), word classes (42%), word forms (42%), run-on sentences (33%), word omissions (25%) and fragments (25%). Similarly, these grammatical errors continued to occur with another grammatical error, subject-verb agreement (19%), but the occurrence rate of grammatical errors more than halved in the post-test (see Table 25).

Pre-test			**Post-test**		
No	**Grammatical errors related to lexical bundles (12 sentences in total)**	**Rate**	**No**	**Grammatical errors related to lexical bundles (32 sentences in total)**	**Rate**
1	Lexical bundle choices	58%	1	Lexical bundle choices	28%
2	Word choices	58%	2	Word classes	25%
3	Word classes	42%	3	Subject-verb agreement	19%
4	Word forms	42%	4	Word choices	13%
5	Run-on sentences	33%	5	Word omissions	9%
6	Word omissions	25%	6	Run-on sentences	9%
7	Fragments	25%	7	Fragments	3%
			8	Word forms	3%
No	**Meanings**	**Rate**	**No**	**Meanings**	**Rate**
1	Meaningful	25**%**	1	Meaningful	75%
2	Direct translation	17%	2	Direct translation	25%

Table 25 Grammatical errors related to the use of the lexical bundles in the pre-test and post-test

The findings revealed that there was more than one grammatical error within the same sentence. In examples 1–12 below, grammatical errors (in bold) seemed to co-occur with the lexical bundles used by the students (underlined). The analyses showed that the longer the sentences that the students produced, the more there was a tendency for students to make more than one grammatical error in a sentence.

Wrong choices of lexical bundles in the pre-test, examples 1–4:

(1) *My project due to the* ***environment of Thailand****.*
(2) *I want study of Kasetsart University, for the development of* ***me****.*
(3) *I want the formation of* ***my family****.*
(4) *I have installed solar cell the number of* ***three kilowatt****.*

The findings showed that there were no correct sentences in the students' writing production (0%). Wrong choices of lexical bundles were seen in 58% in the students' writing production. In term of lexical bundles' structures, examples 2-4 correctly use *the sequence + n + of the* (the development of, the formation of, the number of). However, the students failed to choose the right lexical bundles in the right contexts and were not able to use correct grammatical points in producing complete and correct sentences. For example, sample (4) shows that the student wanted to explain a *3-kilowatt solar cell* rather than *'the number of three kilowatts'*. The lexical bundles in examples 1-4 fall into the 'referential expressions' group (Biber et al., 1999: 1016). Regarding meanings, the students were expected to use the target lexical bundles to refer to number or quantity, e.g, *the number of*, or to describe the process of an event lasting for a period, e.g.,*the development of, the formation of* (§ 3.3 functions of Top 50 LBs in Chapter 3). As a result, the students failed to use correct lexical bundles in the specified contexts; rather they used the lexical bundles given to tell their personal stories. It refers to the way students applied the lexical bundles they learned during the teaching task. Instead of using these bundles in more formal or academic contexts, as intended, students incorporated them into personal narratives or examples from their own experiences. This behavior highlights the students' preference for connecting new language items with their own lives, which may have been influenced by their familiarity with conversational rather than academic writing. This observation

suggests the need for further scaffolding to help students apply lexical bundles in discipline-specific contexts more effectively.

Grammatical errors related to the lexical bundles used in the pre-test, 5–12*:*

(5) *The use of solar energy Instead of using the coal can* ***be*** *the amount of* ***pollutants because many aspects such as the air landscapes.***

(6) *Solar cells for study the development of* ***solar frame but can see how to keep underneath and the*** *formation of* ***solar cells*** *that as a result* ***how, Impressions from*** *the number of server results*

(7) *The use of* ***electricity in school building****.*

(8) *In technology* ***design must have*** *the effect of* ***produce and protected system*** *to best, future the number of* ***many people everyone want have need in technology on intern swell up****.*

(9) *I'm study electrical engineering due to the* ***liked it and*** *the effect of* ***do it well*** *but I'm not good English.*

(10) *Solar energy is heat energy from sunshine manufacture electricity one way to reduce the effect of* ***not Earth and manufacture electricity*** *the performance of* ***is one option****.*

(11) ***The new of my produce is Energy saving lamp*** *in this study* ***can reduce*** *Global warning as well as* ***can reduce*** *electricity charge.*

(12) *But the number of* ***many students****, makes me match with someone students.*

A closer examination of the words following lexical bundles (*the effect of'* and *the number of)* revealed incorrect noun phrase forms, *produce and protected system, many people everyone, many students (*42*%)* used in examples *(*8*)* and *(*12*)*. The words *many students* are grammatically correct to quantify count *(*Quirk et al*.*, 1985; Biber et al*.*, 1999*)*. However, in *(*8*)* and *(*12*)*, the word *many* should not be placed in front of the word *students.*

Consequently, the longer the sentences the students produced, the greater the tendency to produce more incorrect sentences. The second grammatical error found was improper *word choice (*58*%)* by the students in the sentences. In *(*5*)*, the student used an improper verb choice, be, instead of possibly *decrease* in the sentence.

One of the grammatical issues identified in the students' sentence production, accounting for 42%, involved errors in the use or application of word classes. Errors in the use or omission of the verb *be* were observed in the students' writing production.

The absence of the verb *be* can be seen in example *(*1*)*, there is no verb *be* linking a subject and subject complement. A missing gerund, *studying*, and preposition, at, can be found in example *(*9*)*. Once again, the wrong choice of lexical bundle can be found in the same statement. A student may want to mean *because* instead of *due to* and *as a result* instead of *the effect of.*

Word omission occurred in 25*%* of the students*'* writing production, as can be seen in statement *(*11*)*.

The results indicate that 33% of the errors in statement (6) are classified as run-on sentences. This occurs when two or more independent clauses are combined without proper conjunctions or punctuation, leading to grammatical inaccuracies. For example, the pre-test data highlights a notable presence of run-on sentences, accounting for one-third of the total errors. In the post-test, however, the occurrence of run-on sentences decreased to 9%, reflecting an improvement in the students' ability to construct grammatically accurate sentences. This aligns with the overall trend observed in the post-test results, where grammatical accuracy showed significant progress across various categories.

Regarding meanings, the students tried to express more than one idea in the sentences, as can be seen in examples *(*10*)* and *(*3*)*. For example, in statement *(*3*)* the idea is incomplete, but the student thinks it's a correct sentence that has clear semantic value. As a result, fragments *(25%)* appeared in the students' sentences production, as can be seen in samples *(*6*)* and *(*7*)*. The effect of direct translation from the students' L1 is clearly observable in statement (6). The students attempted to connect fragments, resulting in meaningless statements.

Also, morphological errors can be seen in examples *(*2*)*, *(*6*)* and *(*8*)*. Examples *(*2*)* and *(*8*)* reveal the use of repetitive verbs*: want study* and *must have,* which might be the effect of direct translation from the first language *(*Thai*)*. The students may have tried to compensate by using verbs that carry equivalent meanings in Thai and English, resulting in phrases like *want study*. In the Thai language system, there is a subject-verb-object sequence. The verbs 'must have' in *(*8*)* are grammatically correct but are meaningless when combined in this sentence. Further, with more use of direct translation, more persistent

grammatical errors are found in examples *(*2*)* and *(*8*)*. Once again, in example (6), the entire sentence lacks a main verb, resulting in the presence of sentence fragments. The structure of *(*6*)* is also as a result of word-for-word translation, the student attempted to link fragments together without the use of verbs, articles and punctuation. Not all of these word classes appear in Thai structure. Example *(*8*)* shows a subject-verb-object sequence translated directly from Thai. In Thai syntax, sentences can start with a verb *(*as an exclamation or question*)*, or comprise a series of nouns without verbs. The use of ellipsis is normal in Thai so this might be a possible explanation for the production of the sentence in *(*8*)*. Students*'* reliance on L1 equivalence to L2 structure is in line with studies by Sattayatham & Honsa, *(*2007*)*, Kaweera *(*2013*)*, Sukasame, Kantho, and Narrot *(*2014*)* and Watcharapunyawong & Usaha *(*2013*)*. Although their average years of studying English is more than eight before entering university, they have restricted knowledge of English proficiency, in line with studies of learners*'* specific variables by Ädel *(*2015*)* and Cobb and Boulton *(*2015*)*.

In the post-test, there was an improvement in the grammatical errors rate in the students' written sentences. As shown in Table 26, the percentage of grammatical errors found in students' writing production in the post-test dropped to less than a third of that in the pre-test. Similarly, the number of correct sentences produced by the students increased, 12 sentences in the pre-test, 32 sentences in the pos-ttest. The number of correct sentences increased dramatically from 0% in the pre-test to 66% in the post-test, resulting in a decrease in the rate of wrong choices of lexical bundles from 58% in the pre-test to 28% in the post-test.

Similarly, the students could successfully convey meaningful sentences, resulting in threefold increase in the rate of clear meanings in the post-test (from 25% to 75%). For example, in (15), the student uses the lexical bundle 'the performance of' to explain and describe 'the sunlight' lasting for a period correctly. In the same vein, the student tried to make this statement relevant to his/her project. The use of appropriate lexical bundles can also be seen in the following examples: (13), (14), (16), (17), (18), (19) and (20).

Samples of students' statements in the post-test, (13) to (20):

(13) *The effect of* ***green house effect*** *makes it earth warming.*

(14) *The smartphone is used to* ***connect to a wifi****.*

(15) *In order to study the use of Solar cell, I analysis the performance of* ***sunlight*** *in Kasetsart University.*

(16) *In order to* ***provide convenience*** *to a people, I am doing a project on the use of* ***on-off smartphone****.*

(17) *The use of water* ***circulation*** *has good quality.*

(18) *The number of* ***Thai people*** *use electricity is increasing due to the higher temperature.*

(19) *In order to analyse the effect of* ***the sunlight****, I used in the production of* ***electricity in the home****, it* ***help save on*** *energy costs.*

(20) *We can measure PD In order to* ***protect*** *the brack down.*

Subject-verb agreement (SVA) errors accounted for 19% of the grammatical issues identified in the students' writing. Although SVA is not typically considered a complex grammatical error, it appeared to significantly challenge the students. Additionally, errors related to word classes persisted as another notable issue in their writing. However, the rate decreased substantially from 42% to 25%. Taking (19) as an example, the first chunk of the sentence that contains the lexical bundle *'the effect of'* seems to be correct. However, it appears that the second chunk that consists of the lexical bundle *'the production of'* is incorrect regarding grammar (SVO) and meaning. Hence, wrong choices of lexical bundles (e.g.,the production of) still remained in the post-test, but dropping by over half from the pre-test (from 58% in the pre-test to 28% in the post-test).

The following samples (21) and (25) represent both wrong choices of lexical bundles and SVA.

More examples of wrong choices of lexical bundles can be seen in (21), (22), (23) and (24). The lexical bundle choice 'the use of' and the word choice 'have age' were not appropriately used in the sentence. Instead, the chunk '20-30 year' might possibly be preceded by the chunk 'a life span' in (21). Although wrong word choices remained in the post-test, the occurrence rate dropped sharply from 58% to 13%. The wrong choice of word

may apply to the use of the word ‘Current’ as an ‘adverb’ instead of perhaps ‘Currently’ or ‘At present’, in example (27).

Samples of students’ statements in the post-test, (21) to (29):

(21) *The power transformers **have age** the use of **20-30 years**.*

(22) *The number of **friends that is** in a group have skill on solar energy due to the **sunlight is clean energy** and in infinite energy.*

(23) *When is the temperature changes makes voltages Due to the **I did a project temperature changes**.*

(24) *I have a computer to the use of **the search for information**.*

(25) *The tank is Metal and Sealed, The brack down **of transformer** due to the **PD in transformer**.*

(26) *I did a project on energy due to the **Energy used on** much higher.*

(27) ***Current,** the number of **car is increasing** every year cause global warming due to the **carbon dioxide accumutate** in the atmosphere.*

(28) *I had to use plywood in the project, but if the number of **plywood** is very high.*

(29) *OHMASA_GAS a fluid stirrer that does not (missing word) is used to (missing word) **mechanical vibrations** of low frequency instead.*

Examples (21), (23), (24), (25), (28) and (29) that are discussed here are potentially more closely linked to direct translation from L1 (Thai). Once again, in the statements (21) and (29), there is use of the verb ‘have’ and the noun ‘age’ (in 21), and there are double verbs, ‘does not’ and ‘is used to’, in (29). In fact, both instances reveal the pattern of Thai and English structure equivalence, subject-verb-object (SVO) sequence. Students might try to find words in English that have similar meanings in Thai to produce the SVO sequence.

The omission of words is still also found in (25) and (29), though the percentage of occurrence has decreased noticeably from 42% to 9%. The missing main verbs in (25) and (29) suggest the effect of direct translation from L1. Regarding the influence of L1 sentence construction, it eventually made the sentences meaningless, see also examples (28) and (29). These instances appear to be clauses and phrases without any punctuations or prepositions to join the sentence parts.

Similarly, this issue led to the production of numerous run-on sentences (ROS) by the students, as illustrated in examples (23) and (22). Although run-on sentences were present, their occurrence decreased significantly from 33% in the pre-test to 9% in the post-test. Another grammatical error in (23) is fragments, though this decreased significantly from 25% in the pre-test to 3% in the post-test. The use of the wrong form of noun phrases preceded by the lexical bundles 'due to the' in (26) and 'the number of' in (27) still appeared in the post-test, though the rate fell considerably from 42% in the pre-test to 3% in the post-test.

4.3.2.2. Grammatical errors not related to lexical bundles

Grammatical errors not related to the use of the lexical bundles still featured in the students' writing production, as can be seen in Table 26 below. There are eight types of grammatical errors that occurred in the pre-test and the post-test, although each test type had some error types that were not found in the other test type. Grammatical errors related to lexical bundles in Section (4.3.2.2) recurred in the students' writing production, i.e.,word classes, word choices, word omissions, word forms, fragments, subject-verb agreement (SVA), run-on sentences (ROS). Further, the rate of occurrence of these errors seemed to be quite frequent. Other types of errors identified in the students' writing include incorrect use of tenses, word order issues, misuse of the to-infinitive, errors in nominalisation, incorrect sentence structure, and incomplete sentences.

No	Pre-test (17 sentences in total)	Rate	No	Post-test (18 sentences in total)	Rate
1	Correct sentences	29%	1	Correct sentences	56%
2	Word classes	53%	2	Word classes	11%
3	Word choices	35%	3	Word choices	11%
4	Word omissions	29%	4	Word form	11%
5	Fragments	24%	5	To-infinitive	11%
6	Tenses	18%	6	Nominalisation	6%
7	Word forms	18%	7	Word order	6%
8	Subject-verb agreement	18%	8	Wrong sentence structure	6%
9	Run-on sentences	6%	9	Incomplete sentences	6%
	Meanings			**Meanings**	
1	Direct translation	35%	1	Meaningful	56%
2	Meaningful	18%	2	Direct translation	11%
3	Incomplete ideas	6%			

Table 26 Grammatical errors not related to the use of lexical bundles in the pre-test and the post-test

Broadly speaking, there is a marked increase in the number of correct sentences, increasing from 29% in the pre-test to 56% in the post-test.

Meaningless sentences in students' pre-test:

(30) *On of the things that students need to meet the electricity is electrocuted.*

(31) *I hope engineering everyone must design produce your best.*

(32) *At present, Thailand is not widespread Because of the high costs and in the area is limited is used to wind is less.*

As can be seen in (30) – (32) above, in these statements it is difficult to work out the meaning of these sentences. While the sentences may contain elements that suggest some level of meaning—such as references to wind in Thailand or high costs—they are constructed in a way that makes their intended meaning difficult to interpret. This lack of clarity could be attributed to the students' limited ability to organize ideas coherently or to their unfamiliarity with the lexical bundles used in the sentences. The revised observation

highlights the need for additional support in helping students construct meaningful and coherent sentences using the target lexical bundles. There is a tendency for word-for-word translation and the students seem to rely on their personal knowledge, so that the three instances are grammatically incorrect. As a result, the rate of the meaningful sentences is less (18%), and they seem to rely on direct translation from L1 (35%). The use of repetitive verbs in (32) 'is limited' and 'is used to' signals direct translation from Thai, yielding the same results as in the previous section (§ 4.3.2.1).

Word choices/fragments/ prepositions/word omissions/tenses in students' pre-test:

(33) *I project on electricity in building.*

(34) *There used to be many.*

(35) *In a study if wind energy, we can see that.*

(36) *I was studying electrical engineering and computer at Kasetsart University.*

More examples of grammatical errors are word choices, fragments, word classes, word omissions. In example (33), issues with word choice and the omission of the verb 'to be' are evident. Perhaps, the student might want to say *'I'm working on a'... in* (33). There is a 'missing object' and the use of a wrong form 'there+ copula be (existence)' in (34). Examples of incomplete sentences/ ideas or fragments can be found in (34) and (35). In particular, (35) is missing another important part, which is object completion. The use of the wrong tense, 'was', in (36) is also noticeable. It is likely that the students might not be aware when choosing words, and another possible explanation is that there are no tenses used in Thai. So, the use of wrong tenses is commonly seen in Thai students' writing.

In the post-test, there is a noticeable decrease in the rate of grammatical errors. In examples (37) – (39) below, grammatical-error free and meaningful sentence production can be seen. In term of context, these instances are relevant to the engineering project the students were working on.

Samples of correct sentences in the students' post-test:

(37) *The effect of transformer is Discharge.*
(38) *I do energy saving house model projects.*
(39) *Ordinary electrolysis makes O2 and H2 gas.*

Nominalisation/word classes in students' post-test:

(40) *I have done about project changing temperature between heating and cooling.*

Word omission/wrong sentence structure/to-infinitive

(41) *I have to do this one project.*
(42) *This low frequency vibration breaks water surface tension, so bubbles can't occur is electrolysis.*
(43) *I do learning potential electricity production from Solar cell project.*
(44) *It can be automatic following.*
(45) *This robot cannot contain over to 50 kilograms.*

Even though the rate of grammatical errors decreased substantially in the post-test, there were persistent grammatical errors, and other grammatical errors occurred in the post-test. Based on these results, there is a co-occurrence of the persistent grammatical error 'word classes' and other grammatical errors 'nominalisation and word order' in (39). These grammatical errors show the same proportions in the students' writing production. The errors wrong use of nominalization *'changing'*, missing indefinite article *'a'* and wrong word order can be seen in (40) above.

In (41), the use of the word *'one'* to count a singular noun *'project'* should be omitted because in this context the pronoun *'this'* identifies the noun *'project'*. The wrong structure can be seen in (42). The wrong verb form '*do + learning'* is incorrectly used in (43).

Regarding meanings, the use of direct translation can be seen in examples (44) and (45), these two examples do not make sense.

This section addresses grammatical errors related to the use of lexical bundles, and errors not related to lexical bundles. The results revealed that there were grammatical errors associated with lexical bundles, such as word choices, word classes, subject-verb agreements, run-on sentences, fragments and word form (§ 4.3.2.1 and 4.3.2.2).

Grammatical errors associated with lexical bundles appeared in the pre-test and post-test. Even though the rate of grammatical errors dropped dramatically in the post-test, there were occurrences of grammatical errors in the students' writing production.

An association between lexical bundles and grammatical errors in the pre-test and post-test can be identified in two aspects: bonding and burying (followed Osborne, 2003: 81). Grammatical errors (in bold) and lexical bundles (underlined) seem to be bonded to each other. For example, there are wrong choices of lexical bundles in examples (8) and (12) in the pre-test, *'the effect of' produce and protected system'* and *'the number of many students'*. The findings of the research are in line with the study by Osborne (2003). He identified the occurrence of grammatical errors with lexical bundles in advanced level students in two things: bonding and burying. In his study, he found out that

> "bonding occurs when lexical or grammatical elements that have formed associations in the learner's lexicon or grammar may become bonded to each other in such a way that they not only co-occur, appropriately or not but also appear in adjacent position" (ibid.: 81).

Example (19) represents the concept of burying, *'*elements which are embedded inside larger units may become less salient, and so lose grammatical features that they would normally be expected to carry*' (*Osborne, 2003*:* 81*).* As in (19), the first chunk that consists of a bundle is correct *'To analyse the effect of the sunlight, I used in the production of electricity in the home'*. However, as the sentence gets longer, *'it help save on energy costs',* there is a tendency to omit the third person –s in the sentence.

The results suggest that the use of lexical bundles triggered some grammatical errors in the students' sentence production, and this is consistent with previous studies by Osborne (2007) and Paquot and Granger (2012). One primary reason was that students often struggled with the grammatical structure of the lexical bundles themselves. For example, bundles like *in the case of* or *it is important to* require precise integration with the surrounding sentence elements, such as correct prepositions, verb forms, or article usage. Many students misused these bundles by either omitting essential components or

combining them with incorrect structures, resulting in ungrammatical sentences. Another reason was that students frequently misunderstood the contextual appropriateness of certain bundles, leading to their misuse in sentences where they did not fit logically or syntactically. For instance, using *the purpose of this study is* in a context that required a concluding remark created coherence issues and triggered errors in the sentence structure. Additionally, students' limited proficiency in English often compounded these problems, as they struggled to balance the complexities of using lexical bundles with broader grammatical rules. These challenges highlight the importance of targeted instruction and practice to help students master both the form and function of lexical bundles in academic writing.

Regarding meanings in the students' writing production, the students were not successfully able to convey the ideas in the specified topic. It can be seen in the pre-test that the students produced sentences relevant to their background knowledge or experience. The current study is in line with that of Hinkel (2003) who explains that students were prompted to do written activities so they could draw on their experiences, resulting in writing personal narratives or statements of belief/opinion (Hinkel, 2003). It is interesting to note that such a prompted writing activity in the pre-test and post-test may have hindered the students' ability to write. The prompted writing activity used in both the pre-test and post-test may have impacted the students' ability to express themselves freely. This is because the structured nature of the prompts might have limited their creativity and natural flow of ideas, forcing them to conform to specific lexical and grammatical patterns. While this approach was designed to focus on the use of lexical bundles, it might have unintentionally created a sense of constraint for some students, making it harder for them to write more freely or intuitively. Further discussion of this observation is included in Chapter 5, where I reflect on the implications of this finding and provide recommendations for future teaching and testing approaches.

One possible reason is that the prompts provided may have restricted the students' creativity and flexibility in generating their own ideas, leading to a reliance on the specific lexical bundles emphasized in the activity. This limitation might have caused students to

focus excessively on using the prescribed bundles, rather than developing coherent and contextually appropriate paragraphs. Additionally, the pressure to include specific bundles may have diverted their attention from other critical aspects of writing, such as grammar, sentence structure, and overall flow. Another factor could be the cognitive load associated with prompted writing tasks. For students with limited proficiency, simultaneously managing lexical bundles, adhering to the prompt, and maintaining grammatical accuracy may have overwhelmed their processing capacity, resulting in less effective writing performance. Furthermore, the unfamiliarity of the test format might have added stress, hindering their ability to apply their knowledge in a natural and confident manner. These factors suggest that while prompted activities are useful for targeted assessment, they may inadvertently constrain students' ability to demonstrate their full writing potential in less structured contexts. As a result, the students produced sentences relevant to their background knowledge or experience. However, the effect of lexical activities in the pre-test and post-test are not the main focus of this study. Perhaps the grammatical errors that occurred in the students' sentences production were due to their limited lexical repertoire, language proficiency and experience of writing research articles (see Paquot, 2012). The students were engineering students who have mixed ability in English proficiency, little experience in writing English in specific academic prose, and no experience in working with concordances. Based on the teaching of lexical bundles, the grammatical errors found in the students' writing in this study might be different or similar to grammatical errors made by different groups of students.

4.3.2.3. The relation between bundle-specific and general errors

The data illustrates significant changes in the types and frequencies of grammatical errors before and after instruction. In the pre-test, only 29% of sentences were correct, with a high occurrence of errors in word classes (53%), word choices (35%), and fragments (24%), among others. These errors reflect general linguistic difficulties, particularly in constructing syntactically accurate sentences and selecting appropriate words.

In the post-test, the percentage of correct sentences increased dramatically to 56%, while errors related to word classes, word choices, and other grammatical categories significantly decreased (to 11% each). These results suggest that targeted instruction or practice, possibly involving lexical bundles, may have contributed to this improvement by helping students internalise correct syntactic structures and word usage. Regarding meanings, in the pre-test, only 18% of sentences were considered meaningful, while 35% exhibited direct translations and 6% showed incomplete ideas. After instruction, 56% of sentences became meaningful, indicating a marked improvement in students' ability to produce coherent and contextually appropriate sentences. This improvement could be related to their exposure to and practice with lexical bundles, as bundles often encapsulate meaning-rich, formulaic expressions that help convey clear ideas.

The connection to lexical bundles lies in their role as scaffolds for language production. By internalising frequently used patterns (e.g., noun or verb phrase fragments), students may have been able to avoid common errors such as word omissions, incomplete sentences, and fragmented ideas. Additionally, lexical bundles may have reduced errors related to word choices, as they provide ready-made, contextually appropriate phrases, thereby minimising incorrect or literal translations.

However, some residual errors in areas such as nominalisation (6%), word order (6%), and wrong sentence structure (6%) in the post-test suggest that while lexical bundles are effective in improving language accuracy, they cannot fully address deeper grammatical and syntactic issues. These require additional explicit instruction and practice.

In conclusion, the data supports the idea that the use of lexical bundles contributes positively to reducing errors and improving sentence accuracy and meaning, but additional efforts are needed to address broader grammatical issues.

4.3.3. Vocabulary knowledge extension (results from the students' lexical bundles uses in the pre-test and post-test)

As previously mentioned, interviews with the students were carried out after the lexical bundles instruction. The interviews with 13 students were transcribed and translated from

Thai into English. The transcribed data were interpreted and grouped into four main aspects: lexical bundles teaching, AntConc concordances, materials contents and students' attitudes towards English. Information on the interviews, students' notes and teacher's field notes were drawn together to discuss the effectiveness of teaching the workshop.

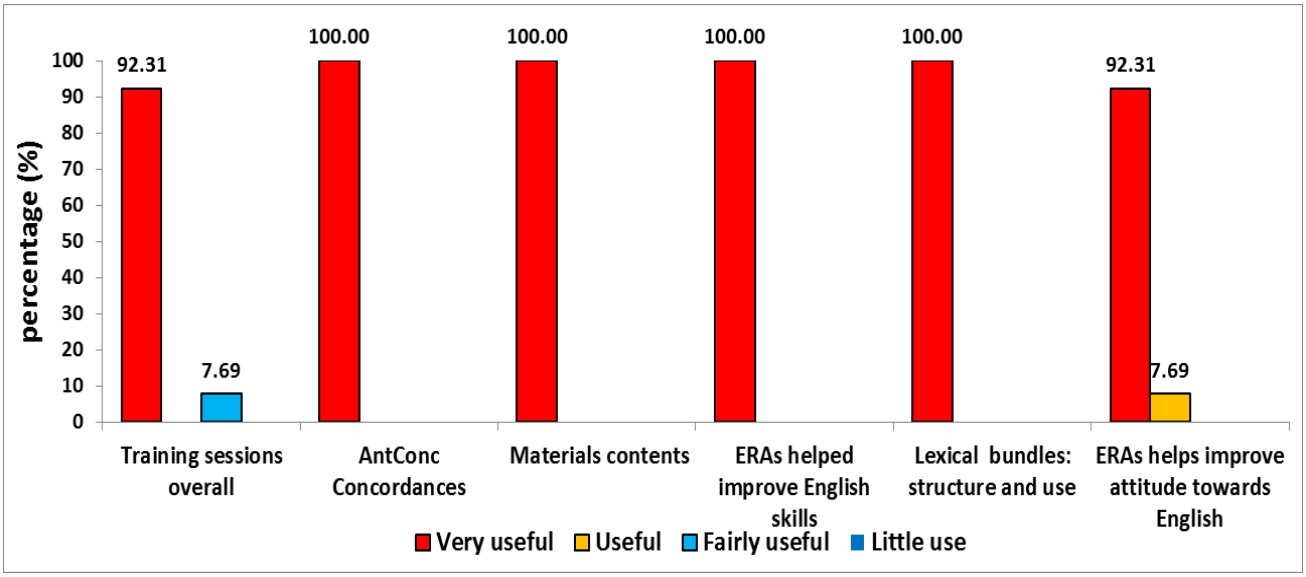


Figure 6 Student evaluations of the workshop from interviews (n = 13)

Overall, 12 students reflected on the training session as having been very useful ***(92.31%)*** and only one out of the 13 regarded the training session as somewhat useful ***(7.69%).*** They agreed that they had gained vocabulary knowledge through learning lexical bundles (see Figure 5).

Concerning teaching lexical bundles (structure and use), the students added that the target bundles were useful (S1) and easier to remember than individual words (S3 and S5), as can be seen from the students' interview extracts below.

> ***'I think it is useful. I think I will have to use a lot of them'. (S1)***
> ***'I feel more comfortable learning bundles than individual words'. (S3)***
> ***'I think chunking is easier to remember such as the bundle 'the number of', I immediately can remember it'. (S5)***

Gray and Biber (2015) write that lexical bundles are extended collocations. While doing the activities, the students not only learnt target bundles but were also taught the words that are commonly used with the bundles. Cf. Firth (1957:11), who stated that 'you shall know a word by the company it keeps'; this study similarly analyses lexical bundles within their contextual collocations. The students could extend their vocabulary, in particular, words that are unique to the engineering field. See for instance the extract from a student interview (S12) below, in which she agrees that she gained vocabulary skills from the workshop.

> *Me: To what extent do you think the ERA sessions helped to improve your English skills?And in what aspects (e.g.,word chunks,grammar, syntax, semantic))?*
> *S12: **My vocabulary**, loads of them in the sessions, structures, patterns of those words*
> *Me: Can you give me **examples of words you learnt** and say to what extent the LBs helped to **improve your vocabulary**?*
> *S12: It was not only LBs that I learnt, I did notice the words that come before or after LBs. I do not normally pay attention to the words that come before or after the LBs. However, I did pay attention to their meanings, and types of words.*
> (Extract from student interview, translated from Thai)

Apart from vocabulary skills, she noticed patterns and the word classes of co-occurring words as well. At the end of the session, it was expected that the students would have been provided with a stock of words that occur in the context of engineering, which can lead them to better L2 proficiency (Martnez & Schmitt, 2015).

4.3.4. Creating corpus-informed materials: lexical bundles and data-driven learning (DDL) as the focus

4.3.4.1. Authenticity of the ERAs corpus

Authenticity is paramount in materials production because it provides context and language features specific to that register (Meunier & Reppen, 2015). Materials were produced by

using 100 research articles from engineering majors (see Chapter 3). Research reports were used in the teaching and learning of engineering students at KU. CSC. The ERAs corpus was built to represent the authentic use of lexical bundles in an engineering context.

4.3.4.2. Materials activities in the classroom

Materials informed by the corpus were divided into two main activities: lexical bundles, and data-driven learning with the use of AntConc when working with bundles in ERAs. Materials were produced and trialled with 11 participants before using them in the classroom (§ 4.2.2.2 in this Chapter). Suggestions elicited from the pilot study were drawn on to revise the materials accordingly. Aspects that needed to be addressed included materials activities, activities order and language use.

After the materials were revised, they were employed in the workshop. Following the common teaching convention in the classroom of Present, Practice, Product **(**PPP, Ur, 1996**)**, there was provision of teaching to present the use of target lexical bundles, practising analysing concordances, and working on activities from the materials provided**.**

Figure 6 shows a sample of teaching the target bundle 'due to the' from concordances, using data-driven learning (DDL) to encourage learner independence (Anthony, 2016). The students were asked to analyse some concordances, five concordances each time, until they finished, and then they shared their answers, working in pairs or as a whole class (see also Sripicharn, 2010; Hunston, 2002).

1	characterised as \xD2easy-to- grind\xD3,	**due to the** absence of hard and abrasive quartz
2	prolonged OSA procedure and inaccurate OSA result	**due to the** absence of customers and limited
3	months), and the rate of occurrence increases	**due to the** acceleration of the degradation process
4	rators. Practical factors including internal tem-	**due to the** action of an open/closed
5	produce a UHPFRC with low binder amount.	**Due to the** addition of fibres, the fibres
6	the cement hydration. On the other hand,	**due to the** addition of limestone and quartz
7	iderable improvement of freeze\xD0thaw resistance	**due to the** addition of SAP has been
8	uniform as the FSCW machine\xD5s	**due to the** additional PMs. This is an
9	may have dramatic decay of adsorption capacity	**due to the** aging process [20].Generally, few i
10	sequences of them, is a challenging part	**due to the** amount of information that can
11	a break down of Si\O bonds	**due to the** attack of alkali hydroxides. Secondly,
12	band at around 1110 cm? 1 in sample CM	**due to the** attack of MgSO4 can induce
13	by DLC1 and DLC2 strategies, respectively;\xA5	**Due to the** better performance, the DLC3 strategy
14	low pH in pore solution and partly	**due to the** blocking effect of brucite on
15	neous deformations are high enough and restrained	**due to the** boundary conditions, stresses can devel
16	uting reduces the power consumption by only 1.5%,	**due to the** buffer leakage power at very
17	in a 1% mass loss at 1100 oC, probably	**due to the** burning off of organic binders.
18	\xDFuid jet tended to separate more readily	**due to the** centrifugal, centripetal and Coriolis f
19	to rotate the mass \xDFow rate increases	**due to the** centrifugal pumping effect. Thus the
20	for the insert at location 1. This was	**due to the** centrifugal pumping effect. By varying

Figure 7 Concordance example of the bundle 'due to the' in Activity C, Session 2.1

They were asked to notice words that co-occurred with the target bundle (3 words from the right). Then, the materials from Session 2**.**1 **(**Activity 2 from Activity C**:** Knowing about engineering actions with the use of lexical bundles**)** were introduced to categorise the groups of words that followed the bundles into groups, categorising the words that were used with target lexical bundles**.**

Figure 6 serves as a more effective example of a concordance for teaching purposes compared to earlier examples. However, it is acknowledged that the length of the concordance lines may not always provide sufficient context for students to fully identify certain patterns, such as *due to the [noun phrase] of*. For instance, line 18 may not contain enough text to determine whether this pattern is present. To address this limitation, future teaching tasks could include extended concordance lines or additional context for each line.

This would help students better understand and analyse the syntactic and semantic patterns within the concordances. These recommendations are highlighted in the conclusions and implications for future materials design.

To give an example of students' involvement in concordancing activities, Figure 7 shows that the students noted down the use of the target lexical bundle 'the number of' in engineering research articles in terms of form, structures and meanings. The words that co-occurred with the target bundles were related to the electrical and computer engineering field**.** For example, students listed co-occurring words relating to the electrical and computer engineering field, such as **'**the use of**'** with committed units, multi physics software, alternative paths **(**see Figure 7 for sample of S8**'**s worksheet**).** Moreover, she identified the word class **(**noun**)** and noted down the meanings of co-occurring words in Thai**.**

In terms of the structure of the target bundles, students noted down the structures and meanings of target bundles that appeared in the concordances **(**see Figure 7 for a sample of S8**'**s notes on the structures of target lexical bundles**).** The target bundles were translated into Thai and the groups of the words that followed the target bundles were categorised into types of materials**.** The words that followed the target bundles were noun phrases and they had specific meanings or statuses in the ERAs corpus**.** The lexical bundles that were used before noun phrases **(**materials in engineering**)** can indicate the status of materials**.** For example, the word **'**absence of…' **(**in 1st and 2nd lines in Figure 6 above**)** can convey the non-existence of materials or non-materials **(**quartz, customers**)** in the ERAs corpus**.** From Figure 7, it can be interpreted that the students not only paid attention to the meanings of target bundles and words that co-occurred, they could also expand their knowledge of vocabulary used in engineering**.** For example, the student **(**S8**)** is able to know that the bundle *the use of* can be used with the lexical word, i.e., *networking cloud* in the computer engineering**.** These findings are in line with vocabulary knowledge extension **(**§ 4.3.3**).** This is a common practice found in the teaching of lexical bundles in the classroom**.**

Item	Groups of Words	Words that Co-occur	Types of Words	Indicate What
1	the number of	committed units	nouns	จำนวนหน่วยที่ถูกมอบหมาย
2	the effect of	percentage	nouns	ผลของเปอร์เซ็นต์
3	the use of	a high	nouns	การใช้ระดับสูง
4	the use of	software	nouns	การใช้ซอฟต์แวร์
5	the use of	networking cloud	nouns	การใช้ระบบคลาวด์เชื่อมโยง
6	the effect of	DG units	nouns	ผลของหน่วย DG
7	the effect of	different connective	nouns	ผลของการเชื่อมต่อที่แตกต่าง
8	the number of	buffers spaces	nouns	จำนวนพื้นที่บัฟเฟอร์
9	the number of	alternative paths	nouns	จำนวนเส้นทางทางเลือก
10	the number of	faulty	nouns	จำนวนที่ผิดพลาด

Figure 8 Sample of student's answer in Activity C (S8)

➢ **the use of** → the use of + adj.
➢ **the effect of** → the effect of + n
➢ **the number of** → the number of + plural noun + singular verb
➢ **due to the** → effect, due to the + n + cause
➢ **the presence of** → the presence of + adj.
➢ **in order to** → formula in order to + V., S+V

Examples:

➢ **the use of** → การใช้ในทางที่แตกต่าง
➢ **the effect of** → ผลกระทบของปัจจัย
➢ **the number of** → จำนวนที่เพิ่มขึ้นของ
➢ **due to the** → เนื่องจากผลกระทบ
➢ **in order to** → เพื่อลดข้อผิดพลาด

Figure 9 Sample of student's notes (S10)

4.3.4.3. Awareness of the use of target bundles

The students learned more about lexical bundles that are used in the electrical engineering field**.** S8 considered that the contents of the workshop were relevant to the meanings of words used in her major subject, electrical engineering**.** She explained that she learned that the word *current* or *kràsĕɛ* in Thai can be used in the context of engineering and has another meaning *the present time* or *pàtcùban* in Thai**.**

> *Me: What aspects of the training sessions do you like most? Why?*
> *W: The content related to engineering, such as current means 'kràsĕɛ' (กระแส), another meaning is pàtcùban (ปัจจุบัน)*
> **(**Extract from student interview S8**)**

4.3.4.4. DDL encourages thinking skills

One of the students revealed that the activities in the materials were thought-provoking for her. She explained that she was told everything in the classroom, but in the workshop she was encouraged to think first before giving an answer (see extract from student's interview below). Good materials should provoke students' thinking in learning (Tomlinson, 2011).

Me: What aspects of the training sessions do you like most (e.g.,learning materials, content, etc.)? Why?
S12: I like the way you taught us. You not only explained the points that I found confusing or could not follow, you also encouraged us to think first. In other words, you did not explain to us straight away, I had to think it through first before I gave an answer. In the classroom, I get used to spoon-feeding, and the teachers just give me the answers right away. I do not have a chance to think. In the sessions, I had to think a lot and do the parts myself. I can remember and understand the contents.

(Extract from student interview, S12)

Moreover, DDL, using the AntConc concordance tool, promoted 'how to' learn the language at their own pace (Anthony, 2017). Student S8 referred to the previous classroom convention she experienced as 'spoon-feeding', in which the teachers gave her all the answers.

One of the students (S2) would like to use AntConc concordances to study English in another context, apart from engineering research articles. Two students (S2, S1) continued that they can use AntConc when writing project papers and in future jobs.

I think this course seems to be another step in learning English in our engineering field. (Extract from student interview, S2)

Me: Do you think you can extend your knowledge of using LBs in the future, based on what you have been taught?

P: Yes, I think I can apply the knowledge I gained in the workshop when writing my project, in particular by using the AntConc program. (Extract from student interview, S1)

Regarding AntConc concordances, the students were comfortable using AntConc software when learning lexical bundles, as can be seen in the extract below.

The software is very convenient; I can use in and out of the classroom, I am thinking of uploading the texts I am interested in in the file and studying the use of

words in the corpus. I can make my own corpus. (Extract from student interview S12)

4.3.4.5. Students gained in self-confidence, which can foster good attitudes towards English

The sample below from S1 shows that the student used to have less confidence in his proficiency. After he completed the activities in the materials, he found that he could improve English skills.

> *M: So in what aspects, do you think the training sessions helped to improve your English skills?*
> *P: I think vocabulary (1), and my attitude towards English; as I previously told you, I am considered to be a 'very poor English user', but I think I have improved my English, such as nouns (1) and specific words (2) used in my engineering field. And (3) structures.*
>
> *(Extracts from student's interview, S1)*

This section has explained how the materials were produced according to the students' need to improve in writing research articles. These materials offered linguistic features called lexical bundles, identifying their structures and functions, how they are used in engineering research articles. The lexical bundles concept is considered to be one of the most important aspects of language learning (Pawley & Syder, 1983; Nattinger & DeCarrico, 1992; Gray & Biber, 2015). The materials provided served the purpose of helping students to become proficient in lexical bundles before moving on to writing. The materials provided aimed to enable students to become proficient in identifying, understanding, and applying lexical bundles in academic writing. Proficiency in this context referred to the ability to recognise lexical bundles in context, analyse their forms and functions using concordance tools, and apply them accurately and appropriately in both guided and free writing tasks. These structured activities supported students in developing the fluency and accuracy needed to use lexical bundles effectively, providing a foundation before progressing to more complex writing tasks. For example, the materials focused on engineering students' needs, involved the students affectively, created specific activities

based on the use of lexical bundles in an authentic context, and included the students' thinking skills (see also Tomlinson, 2011; Meunier and Reppen, 2015).

Having said that, the interview questions did not explicitly prompt students to reflect on the negative aspects of their experiences with the teaching of lexical bundles. This limitation likely resulted in reduced feedback regarding potential challenges or difficulties students encountered during the workshop. For instance, one of the test conditions showed no statistically significant improvement. This outcome may point to underlying issues, such as gaps in the instructional approach, inadequacies in the teaching materials, or difficulties students faced in transferring knowledge effectively to their writing tasks. From a critical perspective, several factors could have contributed to these challenges. Some students may have found the abstract concept of lexical bundles difficult to grasp or felt overwhelmed by the use of corpus tools like AntConc, especially if they had no prior experience with similar software. Additionally, the workshop's limited duration of 10 hours might not have been sufficient to allow for in-depth learning or sustained practice. To address these issues, future studies should include interview questions that specifically focus on the negative aspects of the teaching and learning process. This would help to capture a more comprehensive view of students' experiences, including their struggles, and facilitate the identification of areas for improvement in the teaching of lexical bundles.

The data indicates that teaching outcomes, particularly in improving students' confidence and linguistic skills, were closely related to their initial levels of proficiency. For instance, S1's reflection highlights a positive shift in attitude and self-confidence, as the student transitioned from identifying as a "very poor English user" to recognising improvements in vocabulary, specific domain-related terms, and grammatical structures. This suggests that learners with lower initial proficiency can benefit significantly from targeted instruction, such as the integration of lexical bundles, which provide them with ready-made phrases and structures to support language production.

However, the limited duration of the workshop and the abstract nature of lexical bundles may have posed challenges, especially for less proficient learners. While the materials

effectively addressed specific linguistic needs and provided authentic contexts for engineering students, some students may have struggled with the cognitive demands of understanding and applying abstract linguistic features. For instance, the use of AntConc tools, which require technological literacy, might have been overwhelming for students with no prior exposure to similar software. Additionally, the workshop's 10-hour duration might have been insufficient for less proficient learners to internalise and apply the knowledge effectively.

On the other hand, more proficient learners may have experienced fewer challenges in understanding and applying lexical bundles, allowing them to benefit more fully from the materials and instruction. These learners are likely to have had a stronger foundation in general English skills, enabling them to focus on refining their use of specialised academic language.

To improve outcomes for learners across all proficiency levels, future instructional designs should consider extending the duration of workshops to allow for more sustained practice. Moreover, including explicit scaffolding for less proficient learners—such as simpler explanations of lexical bundles and additional practice with corpus tools—could help bridge gaps in understanding and application. Collecting detailed feedback on challenges encountered during the learning process, as well as tailoring activities to accommodate varying levels of proficiency, would further enhance teaching outcomes and ensure a more inclusive approach to language instruction.

4.4. Summary

This chapter summarises the teaching processes and outcomes of lexical bundles to engineering students, focusing on their application in academic writing. The first step involved selecting lexical bundles to teach, as detailed in Chapter 3. The implementation phase centred on a workshop that integrated materials development and data-driven learning, supported by the use of AntConc. The materials development process included production, piloting, and revising. The students' progress in using lexical bundles was evaluated through a pre-test and post-test. To ensure the study's validity, inter-rater

reliability was tested by three raters who assessed students' work and identified grammatical errors both related and unrelated to lexical bundles. The association between lexical bundles and specific grammatical errors in students' writing revealed their significance in language learning and material development, particularly for specialised groups such as engineering students.

The final stage involved evaluating the study's effectiveness using multiple research instruments, including pre-tests, post-tests, interviews, classroom recordings, and field notes. The results demonstrated the effectiveness of teaching lexical bundles and corpus linguistics techniques in improving engineering students' academic writing skills. Pedagogically, this study highlights the potential of developing tailored learning materials and incorporating data-driven approaches in EAP and ESP classrooms, with a particular focus on lexical bundles. The findings contribute to bridging the gap between students' writing challenges and their technical knowledge, offering an adaptable framework that can be applied to diverse learner groups and proficiency levels.

This chapter has explored the processes and outcomes of teaching lexical bundles to engineering students, highlighting the pedagogical potential of integrating corpus linguistics and data-driven learning into ESP classrooms. The findings underscore the value of tailored materials and targeted teaching methods for addressing specific challenges in academic writing. These results lay the foundation for broader implications, which will be discussed in detail in the final chapter, where the study's conclusions, limitations, and recommendations for future research and practice are presented.

5 Chapter Five: Concluding remarks

5.1 Introduction

This chapter focuses on summarising the main findings of the study and discussing their implications in the context of the research objectives and related theories and approaches. It begins with an overview of the findings from the data analysis chapters, particularly the results of students' pre-tests and post-tests (see Chapter 4), and links these findings to the literature on lexical bundles and the use of corpora in language pedagogy. The chapter also discusses the generic types of lexical bundles (Top 50) (§5.2) and specific types of lexical bundles (Bottom 50) (§5.2) of the study, along with critical commentary on lexical bundle use in engineering research articles (§5.4), challenges in teaching lexical bundles (§5.5), and broader implications for academic writing (§5.6). This study also reveals limitations of the study (§5.7). Finally, it addresses overall strengths and weaknesses, highlights the contributions of this research and suggests potential directions for future studies on lexical bundles and their applications in language teaching (§5.8).

5.2 Generic types of lexical bundles (Top 50)

The analysis of the Top 50 lexical bundles presented in Chapter 3 provides insights into the structural and functional characteristics of frequently used bundles in engineering research articles. These findings highlight the prevalence of NP-based, PP-based, and VP-based bundles, which together account for over 80% of the observed structures in the ERAs corpus. Rather than repeating the statistical summaries from Chapter 3, this section focuses on the implications of these structural patterns for engineering discourse.

NP-based bundles, such as *the use of* and *the effect of*, are pivotal in establishing clarity and coherence in research writing. These bundles often frame research questions, present results, and signal logical relationships, thereby facilitating the reader's understanding of complex concepts. For instance, *the use of* frequently introduces methodological considerations, while *the effect of* signals causal relationships between variables. Such patterns underscore the importance of explicit teaching of these structures to engineering students, enabling them to articulate their findings more effectively. The

functional analysis further reveals that referential and discourse-organising bundles dominate the ERAs corpus. Referential bundles, such as *due to the* and *as shown in*, help writers connect visual or numerical data to their interpretations, supporting the multimodal nature of engineering communication. Discourse-organising bundles, such as *in order to* and *as a result*, are instrumental in structuring arguments, ensuring logical flow, and highlighting cause-effect relationships. These findings suggest that explicit instruction in these functional uses can enhance students' academic writing skills.

5.3 Specific types of lexical bundles (Bottom 50)

The Bottom 50 lexical bundles, as analysed in Chapter 3, provide a window into the specialised language of engineering. These bundles, characterised by their lower frequency and narrower dispersion, often serve specific functions within the discipline. In this section, I interpret the implications of these findings for materials design and pedagogical practice. Incomplete noun-phrases, such as *flow rate of* and *movement of fluid*, account for the majority of the Bottom 50 bundles. These bundles are integral to describing technical processes and phenomena, such as fluid dynamics and material properties, which are central to engineering research. Although these bundles are less frequent than the Top 50, their specialised nature makes them crucial for students aiming to write within their field. For example, *movement of fluid* conveys precise, discipline-specific meaning that general lexical bundles cannot achieve.

The functional analysis of the Bottom 50 bundles reveals a similar trend: referential and discourse-organising functions are predominant. However, these bundles are more context-specific, often tied to particular experimental procedures or engineering applications. For instance, *fluid leaving the* and *flow resistance in* are bundles that describe specific processes or measurements in engineering studies. These findings indicate the potential for integrating Bottom 50 bundles into advanced instructional materials, helping students become familiar with the technical language they are likely to encounter in their academic and professional careers.

While this study did not explicitly test the Bottom 50 bundles, their potential for teaching cannot be dismissed. The Bottom 50 bundles, as analysed in this study, contain domain-specific language that reflects the engineering context, such as expressions describing processes or materials. Incorporating these bundles into teaching materials could enhance students' ability to engage with discipline-specific texts. However, as the Bottom 50 bundles were not tested in the workshop, their pedagogical impact remains speculative. Future research is needed to explore the effectiveness of using these bundles in teaching engineering students, particularly in relation to their familiarity with specialised vocabulary and grammatical structures.

5.4 Critical commentary on lexical bundle use in engineering research articles

This section critically reflects on the patterns of lexical bundle use identified in the ERAs corpus, examining their implications for academic writing and materials development. By integrating findings from Chapters 3 and 4, this commentary highlights the broader significance of these bundles in engineering discourse while addressing gaps and limitations observed during the study.

5.4.1 Structural characteristics and disciplinary specificity

The structural analysis of the Top 50 and Bottom 50 lexical bundles reveals clear trends in engineering texts. While NP-based, PP-based, and VP-based bundles dominate, clausal fragments and hybrid structures are comparatively rare. This structural predominance aligns with the informational and procedural nature of engineering research, where clarity and precision are paramount. However, the limited representation of more complex structures, such as clausal fragments, suggests a potential underutilisation of rhetorical complexity in student writing. This observation raises questions about the extent to which engineering students are exposed to a diverse range of lexical and syntactic patterns in their academic training.

The specialised nature of the Bottom 50 bundles underscores the need for targeted instruction in domain-specific language. For example, bundles like *flow resistance in* and *fluid leaving the* highlight the technical focus required in engineering texts but also pose

challenges for learners unfamiliar with such constructs. These findings suggest that students need explicit support to bridge the gap between general academic language and discipline-specific terminology.

5.4.2 Functional characteristics and pedagogical implications

The functional analysis demonstrates a significant reliance on referential and discourse-organising bundles, which help writers link data, methods, and interpretations. While these functions are essential for constructing coherent and persuasive arguments, the relatively low occurrence of stance bundles highlights an area for improvement. Stance bundles, such as *it can be observed* or *it is important to note*, can help students articulate their evaluations and establish authorial presence in their writing—skills that are critical for producing impactful research articles.

Pedagogically, the emphasis on referential and discourse-organising bundles provides a solid foundation for materials development. However, incorporating activities that encourage the use of stance bundles could enhance students' ability to engage critically with their data and position their research within the broader academic conversation. This could be particularly valuable for Thai engineering students, who often face challenges in expressing critical evaluations due to cultural and linguistic factors.

5.5 Challenges in teaching lexical bundles

The teaching of lexical bundles presents both opportunities and challenges. While the Top 50 bundles are accessible and directly applicable to students' writing, the Bottom 50 bundles are more complex and context-specific, requiring greater linguistic and conceptual understanding. This complexity is compounded by the limited exposure that many students have to corpus-based learning and data-driven approaches.

Additionally, the findings indicate that grammatical errors related to lexical bundle use persist despite explicit instruction. This suggests that while students can recognise and understand the functional roles of lexical bundles, they require more time and practice to integrate them effectively into their writing. A longitudinal approach to teaching lexical bundles, combined with scaffolded support, may help address these issues.

5.6 Broader implications for academic writing

The patterns of lexical bundle use observed in the ERAs corpus have broader implications for academic writing in engineering. The frequent use of referential and discourse-organising bundles reflects the importance of clarity and logical structure in this discipline. However, the underrepresentation of stance bundles points to a potential area for development in engineering education, where critical thinking and argumentation skills could be more explicitly cultivated.

Overall, this section argues for a balanced approach to teaching lexical bundles, one that combines foundational instruction in general academic language with targeted support for discipline-specific terminology. By addressing the structural and functional characteristics of lexical bundles critically, this study contributes to a deeper understanding of their role in engineering discourse and provides practical insights for improving academic writing instruction.

5.6.1 Expanding the role of lexical bundles in engineering research writing: contributions and distinctions

This study significantly advances the understanding of lexical bundles in engineering academic writing by addressing overlooked aspects in previous research and emphasizing their pedagogical value. Mudraya's (2006) work primarily focused on identifying technical and sub-technical vocabulary from engineering textbooks, but it did not explore multi-word expressions that contribute to structural and functional coherence in writing. This research builds on Mudraya's foundation by shifting the focus from isolated vocabulary items to the dynamic use of lexical bundles in engineering research articles, specifically analysing their roles in constructing arguments, framing research questions, and presenting results. Moreover, while Mudraya's findings informed vocabulary instruction in technical fields, this study goes further by integrating lexical bundles into tailored instructional materials designed to enhance the academic writing skills of Thai engineering students, addressing the critical need for genre-specific writing proficiency.

This study also extends and refines the insights of Cheung (2010) and Graham (2014) by focusing on the nuanced applications of lexical bundles in research articles.

While Cheung's analysis of two-word bundles in the Hong Kong Engineering Corpus highlighted the importance of vocabulary awareness, this research provides a more comprehensive perspective by investigating three-word and four-word bundles, their structural diversity, and their discourse functions. Similarly, Graham's statistically grounded Engineering Phrases List (EPL) identified frequent formulaic phrases but lacked focus on their academic genre applications. By examining how lexical bundles enhance readability, argumentation, and logical flow in engineering research articles, this study bridges these gaps and introduces innovative, corpus-driven teaching materials that address the specific needs of engineering students. These contributions emphasise the value of integrating data-driven approaches to teaching writing, particularly for non-native English speakers, making this study a vital addition to the existing body of research in engineering discourse.

5.7 Strengths and limitations of the study

This section reflects on the overall strengths and limitations of the study. One of the strengths is the novel integration of lexical bundle analysis with the data-driven learning approach, which has not been widely applied to teaching engineering students in Thailand. The findings provide practical insights into how specialised corpora can be used to support language learning. However, the study also faced challenges, such as a small sample size and students' limited familiarity with corpus-based learning tools. These challenges highlight the need for longer-term studies and collaboration between language instructors and subject specialists to enhance pedagogical outcomes. Furthermore, this study offers originality in its focus on engineering-specific lexical bundles and their potential application in ESP contexts.

One of the key weaknesses of this study is the variation in students' ability levels. The participants ranged from beginner to intermediate levels, making it challenging to determine the impact of teaching lexical bundles on specific proficiency levels. This heterogeneity also complicates identifying the optimal stage at which lexical bundles should be introduced in the learning process. Furthermore, the small sample size (n=13)

limits the ability to make meaningful comparisons between students of different levels and hinders generalisations about the findings. Future research should aim to include a larger and more homogeneous sample to better understand the relationship between students' ability levels and the effectiveness of teaching lexical bundles.

Another limitation of this study is the narrow focus on the Top 50 frequent bundles, which, while representative of common patterns in engineering research articles, may not capture the specificity of engineering language. Without incorporating less frequent bundles, such as those in the Bottom 50, the materials may lack depth in addressing engineering-specific discourse. Moreover, the absence of a comparison with other academic corpora limits the ability to confirm whether the bundles taught are unique to engineering or are general academic language features. Future research should explore both frequent and less frequent bundles, alongside cross-disciplinary corpus comparisons, to better contextualise the teaching of lexical bundles and their relevance to engineering-specific writing.

Another limitation of this study is the timing of the post-test, which was conducted shortly after the teaching sessions. While the results indicate an improvement in students' use of lexical bundles, it is unclear whether these gains would be retained over a longer period. Without a delayed post-test, it is difficult to assess the long-term impact of the teaching intervention. Future research should incorporate follow-up studies or delayed post-tests to evaluate the retention of improvements in lexical bundle use and to better understand how these skills are internalised and applied over time.

Another limitation of this study was the similarity of the feedback questions, which may have resulted in students focusing primarily on the positive aspects of the teaching. Future studies should incorporate more diverse feedback questions, including those that explicitly invite students to discuss challenges or areas they found less effective. This approach would provide a more comprehensive evaluation of the teaching materials and methods.

5.8 Future research

Future research could extend this study by exploring related linguistic phenomena such as frames (e.g., the _ of), which are recurrent patterns that often function similarly to lexical bundles. These frames could provide additional insights into discipline-specific language use and support further development of teaching materials.

Another potential avenue for future research is to use more authentic writing samples, such as student essays for coursework, for pre- and post-teaching evaluation. This approach would provide richer data on students' academic writing skills and their application of lexical bundles, as opposed to relying solely on paragraph writing tasks, which sometimes resulted in personal writing not aligned with the target academic genre.

Finally, to better identify engineering-specific bundles, future research should compare the engineering research article corpus used in this study with a more general academic writing corpus. Such a comparison would help distinguish bundles that are unique to engineering from those that are common across academic disciplines, providing a clearer understanding of their specificity and relevance for teaching engineering students.

References

Ädel, A. (2010). Using Corpora to Teach Academic Writing: Challenges for the Direct Approach. In M. C. Campoy-Cubillo, B. Bellés-Fortuño, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. xvii, 281 p.). London: Continuum.

Ädel, 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*(2), 81-92. https://doi.org/10.1016/j.esp.2011.08.004

Aijmer, K. (2009). "So er I just sort I dunno I think it's just because…": A corpus study of I don't know and dunno in learners' spoken English In A. H. Jucker, D. Schreier, M. Hundt, & MyiLibrary (Eds.), Corpora pragmatics and discourse : papers from the 29th International Conference on English Language Research on Computerized Corpora (ICAME 29), Ascona, Switzerland, 14-18 May 2008. Amsterdam; New York: Rodopi.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge [England]; New York, NY, USA: Cambridge University Press.

Altenberg, B. (1998). On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. In A. P. Cowie (Ed.), *Phraseology: theory, analysis, and applications*. Oxford [England] New York: Clarendon Press; Oxford University Press.

Anthony, L. (2011). AntConc 3.2.4 (Version AntConc 3.2.4) [software]. Tokyo: Waseda University.

Anthony, L. (2017). Introducing corpora and corpus tools into the technical writing classroom through Data-Driven Learning (DDL). In J. Flowerdew & T. Costley (Eds.), *Discipline-specific writing: theory into practice*. Oxon; New York: Routledge.

Arnold, I. V. (1986). *The English Word* (2nd Edition ed.). Moscow: Vyssaja Skola.

Aston, G. (1997). Enriching the Learning Environment: Corpora in ELT. In A. Wichmann,

S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora*. London; New York Longman.

Aston, G. (2000). Corpora and language teaching. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora* (pp. 261 p.). Frankfurt am Main; New York: Peter Lang.

Aston, G. (2015). Learning phraseology from speech corpora. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam/Philadelphia: John Benjamins.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baker, P. (2010). Corpus Methods in Linguistics. In L. Litosseliti (Ed.), *Research Methods in Linguistics*. London; New York: Continuum.

Baker, P., & Egbert, J. (2016). *Triangulating methodological approaches in corpus linguistic research*. In *Routledge advances in corpus linguistics 17*. Retrieved from http://www.tandfebooks.com/isbn/9781315724812

Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Bal, B. (2010). *Analysis of Four-word Lexical Bundles in Published Resesarch Articles Written by Turkish Scholars.* (Master of Art Education Master), Georgia State University, Georgia State

Basturkmen, H. (2017). Developing writing courses for specific academic purposes In J. Flowerdew & T. Costley (Eds.), *Discipline-specific writing: theory into practice* (pp. 1 online resource (232 pages)). Oxon; New York: Routledge.

Becker, J. D. (1975). *The phrasal lexicon*. Paper presented at the Proceedings of the 1975 workshop on Theoretical issues in natural language processing, Cambridge, Massachusetts.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. C. Marko (Eds.), *Teaching and learning by doing corpus analysis : proceedings of the Fourth Int. Conference on Teaching & Language Corpora, Graz, 19-24 July,*

2000 (pp. vii, 390 p.). Amsterdam New York: Rodopi.

Bernardini, S. (2004). Corpora in classroom: An overview and some reflections on future developments. In J. M. Sinclair (Ed.), *How to use corpora in language teaching*. Amsterdam John Benjamins.

Biber, D. (2006). *University language: a corpus-based study of spoken and written registers*. Amsterdam; Philadelphia, Penn.: J. Benjamins.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311. DOI:10.1075/ijcl.14.3.08bib

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

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. H. a. S. Oksefjell (Ed.), *Out of corpora: Studies in honor of Stig Johansson*. Amsterdam Rodopi

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In G. N. Leech, T. McEnery, P. Rayson, & A. Wilson (Eds.), *Corpus linguistics by the Lune: a festschrift for Geoffrey Leech* (pp. 305 p.). Frankfurt am Main: Peter Lang.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics, 25*(3), 371-405. DOI:10.1093/applin/25.3.371

Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based Approaches to Issues in Applied Linguistics. *Applied Linguistics, 15*(2), 169-189. https://doi.org/10.1093/applin/15.2.169

Biber, D., Conrad, S., & Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press.

Biber, D., Finegan, E., Johansson, S., Conrad, S., & Leech, G. (1999). *Longman grammar of spoken and written English*. New York: Longman.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2-20. DOI:10.1016/j.jeap.2010.01.001

Boulton, A. (2017). Data-driven learning and language pedagogy. *Language, education and technology: Encyclopedia of language and education*, *3*, 181-192. August 2017. DOI:10.1007/978-3-319-02237-6_15

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL, 17*(1), 47-64. DOI:10.1017/S0958344005000510

Braun, S. (2006). ELISA: A Pedagogically enriched corpus for language learning purposes. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: new resources, new tools, new methods*. Frankfurt: Peter Lang.

Braun, S., Kohn, K., & Mukherjee, J. (2006). *Corpus technology and language pedagogy: new resources, new tools, new methods*. Frankfurt: Peter Lang. Chambers, A. (2005). Integrating Corpus Consultation in Language Studies. *Language Learning & Technology, 9*(2), 111-125.

Brett, D. F. (2010). Identifying chains in spoken Academic English. *Annali della Facoltà di Lingue e Letterature Straniere di Sassari*, *7*, 71-87.

Butzkamm, W. (2009). The bilingual reform: A paradigm shift in foreign language teaching. *Narr Francke Attempto Verlag*. https://doi.org/10.1093/elt/ccr084

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes, 6*(4), 289-302. doi: https://doi.org/10.1016/j.jeap.2007.09.009.

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30-40.

Charles, M. (2015). Same task, different corpus. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning* (pp. 129–154). Amsterdam: John Benjamins.

Chen, H.-J. H. (2011). Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning, 24*(1), 59-

76. doi:10.1080/09588221.2010.526945

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30–49. DOI:10.64152/10125/44213

Cheng, W. (2010). Hong Kong Engineering Corpus: Empowering Professionals-In-Training To Learn Language Of Their Profession. In M. C. Campoy-Cubillo, B. B. Fortuno, & M. L. G. Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. xvii, 281 p.). London; New York Continuum.

Cheng, W. (2010). What Can a Corpus Tell Us About Language Teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics*. London, New York: Routledge.

Choi, B. N. (2015). Lexical Bundles in Economic Research Articles by Native and Non-native Speakers of English. *영학논집 (English Studies)*, *35*, 159-187.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.

Conrad, S. (2017). *From a Plate of Spaghetti to a Cable-stayed Bridge: Increasing the Impact of Corpus Linguistics in Disciplinary Education*. Paper presented at the The 9th International Corpus Linguistics Conference took place from Monday 24 to Friday 28 July at the University of Birmingham, University of Birmingham, UK.

Conrad, S. (2017). The Use of Passives and Impersonal Style in Civil Engineering Writing. *Journal of Business and Technical Communication, 0*(0), 1050651917729864. DOI:10.1177/1050651917729864

Conrad, S., & Biber, D. (2005). The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Lexicographica (2004), 20*, 56-71. https://doi.org/10.1515/9783484604674.56

Conrad, S. M. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education, 8*(3), 299-326. https://doi.org/10.1016/S0898-5898(96)90025-X

Cortes, V. (2002). Lexical bundles in Freshman composition. In R. Reppen, S. M. Fitzmaurice, & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins.

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

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*(4), 391-406. DOI:10.1016/j.linged.2007.02.001

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 3*(1), 43-57. DOI:10.3366/E1749503208000063

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes, 12*(1), 33-43. DOI:10.1016/j.jeap.2012.11.002

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In V. Cortes & E. Csomay (Eds.), *Corpus-based Research in Applied Linguistics: Studies in Honor of Doug Biber*. Amsterdam: John Benjamins.

Cowie, A. P. (1998). *Phraseology: theory, analysis, and applications*. Oxford [England] New York: Clarendon Press; Oxford University Press.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34*(2), 213-238. DOI:10.2307/3587951

Coxhead, A. (2011). The Academic Word List 10 Years On: Research and Teaching Implications. *TESOL Quarterly, 45*(2), 355-362. DOI:10.5054/tq.2011.254528

Coxhead, A. (2015). Acquiring academic and disciplinary vocabulary. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* Hoboken, N. J. : Wiley.

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. [A pilot science-specific word list]. *Revue française de linguistique appliquée, XII*(2), 65-78.

Creswell, J. W. (2013). *Qualitative inquiry & research design: choosing among five approaches* (3rd ed.). Los Angeles, Calif.; London: SAGE Publications.

Csomay, E. (2013). Lexical Bundles in Discourse Structure: A Corpus-Based Study of

Classroom Discourse. *Applied Linguistics, 34*(3), 369-388. DOI:10.1093/applin/ams045

Cunningham, G., Bell, J., Clementson, T., & Redston, C. (2013). *face2face Advanced Student's Book* Cambridge: Cambridge University Press.

Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes, 25*(Supplement C), 71-83. DOI:10.1016/j.jeap.2016.11.005

Dat, B. (2008). ELT materials used in Southeast Asia. In B. Tomlinson (Ed.), *English language learning materials: a critical review* (pp. x, 333 p.). London: Continuum.

Dwight, B. (1976). Meaning and memory. *Forum Linguisticum, 1*, 1-14.

Eckerdal, J. R., & Hagström, C. (2017). Qualitative questionnaires as a method for information studies research.

Ellis, R. (2005). Principles of instructed language learning. *System*, *33*(2), 209-224.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, *20*(1), 29-62.

Eeg-Olofsson, M., & Altenberg, B. (1994). Discontinuous recurrent word combinations in the London–Lund Corpus. In U. Fries, G. Tottie, & P. Schneider (Eds.), *Creating and using English language corpora: Papers from the fourteenth international conference on English language research on computerized corpora.*

(pp. 203 p.). Amsterdam; Atlanta, GA: Rodopi.

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

Firth, J. R., & Association, E. (1951). *Modes of Meaning*: Oxford University Press.

Fligelstone, S. (1993). Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal 17, 17*, 97-110.

Flowerdew, J. (2012). *Discourse in English language education*. In.

Flowerdew, J., & Wang, H. S. (2017). Teaching English for research publication purposes with a focus on genre, register, textual mentors and language re-use: a case study. In J. Flowerdew & T. Costley (Eds.), *Discipline-specific writing: theory into*

practice. Oxon; New York; Routledge.

Flowerdew, L. (2010). Using corpora for writing instruction. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (1st ed.). London; New York, NY: Routledge.

Flowerdew, L. (2012). *Corpora and language education*. In *Research and practice in applied linguistics*. Retrieved from http://lib.myilibrary.com?id=384434

Garner, J. R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research, 2*(1), 31–67.

Garner, J. (2022). The cross-sectional development of verb–noun collocations as constructions in L2 writing. *International Review of Applied Linguistics in Language Teaching*, *60*(3), 909-935.

Gavioli, L. (2014). Exploring texts through the concordancer: guiding the learner. In A. Wichmann & S. Fligelstone (Eds.), *Teaching and Language Corpora*. London, New York: Taylor & Francis.

Gavioli, L., & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT Journal, 55*(3), 238-246. doi:10.1093/elt/55.3.238

Gilquin, G. (2012). Lexical infelicity in English causative constructions. Comparing native and learner collostructions. *Analytical causatives. From 'give'and 'come'to 'let'and 'make*, 41-63.

Gledhill, C. (2011). The 'lexicogrammar' approach to analysing phraseology and collocation in ESP texts. *ASp, 59*, 5-23.

Götz, S., & Mukherjee, J. (2006). Evaluation of data-driven learning in university teaching: a project report. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: new resources, new tools, new methods*

Frankfurt: Peter Lang.

Graham, D. (2014). *The engineering phrases list: Towards teachable ESP phrases.* Paper presented at the 21st Century English Language Education: Towards Global Citizenship, Chiang Mai, Thailand.

Granger, S. (1998). Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In A. P. Cowie (Ed.), *Phraseology: theory, analysis, and applications*. Oxford [England] New York: Clarendon Press; Oxford University Press

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective*. Amsterdam John Benjamins.

Gray, B. (2016). Lexical Bundles. In P. Baker & J. Egbert (Eds.), *Triangulating methodological approaches in corpus linguistic research* New York; London Routledge.

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics, 18*(1), 109-136. doi:doi:10.1075/ijcl.18.1.08gra

Gray, B., & Biber, D. (2015). Phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge, UK: Cambridge University Press.

Gries, S. T. (2008). Phraseology and linguistic theory. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective*. Amsterdam John Benjamins.

Grigaliūnienė, J., & Juknevičienė, R. (2013). Recurrent formulaic sequences in the speech and writing of the Lithuanian learners of English. In S. Granger, G. t. Gilquin, & F. Meunier (Eds.), *Twenty years of learner corpus research: looking back, moving ahead: proceedings of the First Learner Corpus Research Conference (LCR 2011)* (pp. 211-222). Louvain-la-Neuve: Presses universitaires de Louvain.

Gülich, E. (2008). Le recours au préformé: une ressource dans l'interaction conversationnelle. In *Congrès Mondial de Linguistique Française* (p. 072). EDP Sciences.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24*(2), 287-297. doi:10.1111/j.1467-1770.1974.tb00509.x

Hall, G., & Cook, G. (2012). Own-language use in language teaching and

learning. *Language teaching*, *45*(3), 271-308.

Henry, A., & Roseberry, L. R. (2001). Using a small corpus to obtain data for teaching a genre. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT: theory and practice* (pp. xxiii, 419 p.). Amsterdam; Philadelphia, Pa.: J. Benjamins Pub. Co.

Hill, J., Lewis, M., & Lewis, M. (2000). Classroom strategies, activities, and exercises. In M. Lewis (Ed.), *Teaching collocation: further developments in the lexical approach* (pp. 245 p.). London: Heinle Cengage Learning.

Hinkel, E. (2003). *Second Language Writers' Text: Linguistic and Rhetorical Features*. New Jersey: Routledge.

Howarth, P. (1998). The Phraseology of Learners' Academic Writing. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications*. New York: OUP Oxford.

Howarth, P. A. (1996). *Phraseology in English Academic Writing: Some implications for language learning and dictionary making*. Germany: Walter de Gruyter.

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research, 17*(4), 454-484. doi:10.1177/1362168813494121

Hunston, S. (2002). *Corpora in applied linguistics*: Cambridge U.P.

Hunston, S. (2015). Lexical grammar. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge, UK: Cambridge University Press.

Hyland, K. (2000). *Disciplinary discourses: social interactions in academic writing*. New York: Longman.

Hyland, K. (2004). *Disciplinary discourses: social interactions in academic writing* (Michigan classics ed.). Ann Arbor: University of Michigan Press.

Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62. doi:10.1111/j.1473-4192.2008.00178.x

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for*

Specific Purposes, 27(1), 4-21. doi: https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2015). Corpora and written academic English. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge, UK

Cambridge University Press.

Hyland, K., & Tse, P. (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly, 41*(2), 235-253. doi:10.1002/j.1545-7249.2007.tb00058.x

Jalilifar, A., Ghoreishi, S. M., & Emam Roodband, S. A. (2016). Developing an inventory of core lexical bundles in English research articles: A cross-disciplinary corpus-based study. *Journal of World Languages*, *3*(3), 184-203.

John, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. x, 340 p.). Cambridge; New York, NY, USA: Cambridge University Press.

John, T. (2002). Data-driven learning: the perpetual challenge In B. Kettemann & G. C. Marko (Eds.), *Teaching and learning by doing corpus analysis: proceedings of the Fourth Int. Conference on Teaching & Language Corpora, Graz, 19-24 July, 2000* (pp. vii, 390 p.). Amsterdam New York: Rodopi.

Johns, T., & King, P. (1991). *Classroom Concordancing*: Centre for English Language Studies, the University of Birmingham.

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences: acquisition, processing, and use*

(pp. ix, 303 p.). Amsterdam; Philadelphia: John Benjamins Pub.

Kanoksilapatham, B. (2012). Research Article Structure of Research Article Introductions in Three

Engineering Subdisciplines. *IEEE Transactions on Professional Communication, 55*(4), 294-309.

Kazemi, M., Katiraei, S., & Rasekh, A. E. (2014). The Impact of Teaching Lexical Bundles on Improving Iranian EFL Students' Writing Skill. *Procedia - Social and*

Behavioral Sciences, 98 (Supplement C), 864-869. doi: https://doi.org/10.1016/j.sbspro.2014.03.493

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

Koteková, D. (2013). Confidence mastery as the fundamental task in learning a foreign language. *Journal on Efficiency and Responsibility in Education and Science*, *6*(2), 84-104.

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2017). Exploring the Role of Phraseological Knowledge in Foreign Language Reading. *Applied Linguistics, 38*(6), 848-870. doi:10.1093/applin/amv070

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialised corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56-75.

Leech, G. (1997). Teaching and Language Corpora: A Convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora*. London; New York: Longman.

Leech, G. (2015). Descriptive grammar. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge, UK: Cambridge University Press.

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes, 22*(Supplement C), 42-53. doi: https://doi.org/10.1016/j.jeap.2016.01.008

Leńko-Szymańska, A. (2003). Lexical problem areas in the advanced learners corpus of written data. In B. Lewandowska-Tomaszczyk (Ed.), *PALC 2001: practical applications in language corpora*. Frankfurt: P. Lang.

Lillis, T. M., & Curry, M. J. (2010). *Academic Writing in a Global Context: The Politics and Practices of Publishing in English*: Routledge.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31*(1), 25-35. doi: https://doi.org/10.1016/j.esp.2011.07.002

Mackey, A., & Gass, S. M. (2015). *Second Language Research Methodology and Design*. In. Retrieved from http://lancaster.eblib.com/patron/FullRecord.aspx?p=2126959

Maley, A. (2013). Vocabulary. In B. Tomlinson (Ed.), *Applied linguistics and materials development* (pp. vii, 335 p.). London: Bloomsbury.

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*(3), 183-198. doi: https://doi.org/10.1016/j.esp.2009.04.003

Martinez, R., & Schmitt, N. (2015). Vocabualry. In D. Biber & R. Reppen (Eds.), *Cambridge handbooks in language and linguistics* (pp. xviii, 623 pages). Cambridge. Cambridge University Press.

Masuhara, H. (2010). Applications of the Research Results for Materials Development. In B. Tomlinson & H. Masuhara (Eds.), *Research for materials development in language learning evidence for best practice*. London: Continuum.

Maswana, S., Kanamaru, T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand, 2*, 1-11. doi: https://doi.org/10.1016/j.amper.2014.12.002

McEnery, T., Baker, P., & Wilson, A. (1995). A Statistical Analysis of Corpus Based Computer vs Traditional Human Teaching Methods of Part of Speech Analysis. *Computer Assisted Language Learning, 8*(2-3), 259-274.

McEnery, T., & Gabrielatos, C. (2008). English Corpus Linguistics. In *The Handbook of English Linguistics* (pp. 33-71): Blackwell Publishing.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: an introduction* (2nd ed.):

Edinburgh U.P.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. London; New York: Routledge.

Mel'čuk, I. (1998). Collocations and Lexical Functions. In A. P. Cowie (Ed.), *Phraseology: theory, analysis, and applications*. Oxford [England]: New York: Clarendon Press; Oxford University Press

Meunier, F., & Granger, S. (2007). *Phraseology in foreign language learning and teaching*. In. Retrieved from http://lancaster.eblib.com/patron/FullRecord.aspx?p=622977

Meunier, F., Granger, S., Gilquin, G. G., & Paquot, M. (2011). *A taste for corpora in honour of Sylviane Granger*. In *Studies in corpus linguistics (SCL) v. 45*. Retrieved from http://ezproxy.lancs.ac.uk/login?url=http://lib.myilibrary.com?id=323449

Meunier, F., & Reppen, R. (2015). Corpus versus non-corpus-informed pedagogical materials: grammar as the focus In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. xviii, 623 pages). Cambridge Cambridge University Press.

Moon, R. (1998). *Fixed expressions and idioms in English: a corpus-based approach*. Oxford

New York: Clarendon Press; Oxford University Press.

Moon, R. (2009). Sinclair, lexicography, and the Cobuild Project: The application of theory. In R. Moon (Ed.), *Words, grammar, text: revisiting the work of John Sinclair* (pp. viii, 124 p.). Amsterdam: John Benjamins Pub. Co.

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

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the arts - and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: new resources, new tools, new methods*. Frankfurt: P. Lang.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Nattinger, J. R., & Decarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford [England]; New York: Oxford University Press.

Nesi, H. (2015). ESP and Corpus Studies. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* Hoboken, N. J. : Wiley.

Nesi, H. (2016). Corpus studies in EAP. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes*

(pp. 645 pages). London: Routledge.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 283-304. doi:doi:10.1075/ijcl.11.3.04nes

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 283-304. doi:doi:10.1075/ijcl.11.3.04nes

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics, 24*(2), 223-242. doi:10.1093/applin/24.2.223

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. M. Sinclair (Ed.), *How to use corpora in language teaching*. Amsterdam John Benjamins.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam; Philadelphia: John Benjamins.

O'Donnell, M. B., Römer, U., & Ellis, N. C. (2015). The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm. In S. Hoffmann, B. Fischer-Starcke, & A. Sand (Eds.), *Current Issues in Phraseology* Amsterdam / Philadelphia: John Benjamins Publishing Company.

Orr, T. (2006). Introduction to the Special Issue: Insights From Corpus Linguistics for Professional Communication. *IEEE Transactions on Professional Communication, 49*(3), 213 - 216 doi:10.1109/TPC.2006.880750

Osborne, J. (2003). Intergrating corpora into a language-learning syllabus. In B. Lewandowska-Tomaszczyk (Ed.), *PALC 2001: practical applications in language corpora*. Frankfurt: P. Lang.

Osborne, J. (2008). Phraseology effects as a trigger for errors in L2 English: The case of more advanced learners. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 67–83). Amsterdam/Philadelphia: John Benjamins.

Özdemir, N. Ö. (2014). Using corpus data to teach collocations in medical English. *Journal of Second Language Teaching & Research*, *3*(1), 37-52.

Paquot, M. (2010). *Academic vocabulary in learner writing: from extraction to analysis*. London: Continuum.

Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics, 9*(1), 131-156. doi:doi:10.1075/ijcl.9.1.07par

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication*. London: Longman.

Prodromou, L. (2008). *English as a lingua franca: a corpus-based analysis*. London: Continuum.

Qi-bo, Z. (1989). A quantitative look at the Guangzhou Petmleum English Corpus. *ICAME Journal: International Computer Archive of Modern English, 13*, 28-38.

Quirk, R. (1972). *A Grammar of contemporary English*. [London],: Longman.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In J. Svartvik, K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics: studies in honour of Jan Svartvik* (pp. xi, 338 p.). London; New York: Longman.

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

Richards, J. C. (2015). Materials Design in Language Teacher Education: An Example from Southeast Asia. In T. S. C. Farrell (Ed.), *International Perspectives on English*

Language Teacher Education: Innovations from the Field (pp. 90-106). London: Palgrave Macmillan UK.

Richards, J. C., Hull, J., & Proctor, S. (1997). *New Interchange Level 2 Student's Book 2: English for International Communication*: 홍익에프엘티.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*, 141–163.

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction 3, 3*(1).

Salazar, D. (2011). *Lexical bundles in native and non-native scientific writing: applying a corpus-based study to language teaching*.

Salem, A. (1987). *Pratique des segments répétés*. Paris: Publications de L'InaLF.

Samraj, B. (2013). Form and function of citations in discussion sections of master's theses and research articles. *Journal of English for Academic Purposes, 12*(4), 299-310. doi:https://doi.org/10.1016/j.jeap.2013.09.001

Samraj, B. (2016). Research articles. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. xxiii, 645 pages). Abingdon, Oxon Routledge.

Schnmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: acquisition, processing, and use*. Amsterdam; Philadelphia: John Benjamins Pub.

Scott, M., & Tribble, C. (2006). *Textual patterns: key words and corpus analysis in language education*. Amsterdam: John Benjamins Pub.

Simpson, R., & Mendis, D. (2003). A Corpus-Based Study of Idioms in Academic Speech. *TESOL Quarterly, 37*(3), 419-441. doi: 10.2307/3588398

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics, 31*(4), 487-512. doi:10.1093/applin/amp058

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*: Oxford University Press.

Sinclair, J. (2008). The phrase, the whole phrase, and nothing but the phrase. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective* (pp. xxviii, 422 p.). Amsterdam: John Benjamins.

Sinclair, J. M. (1987). *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: Collins ELT.

Sripichan, P. (2010). Designing Corpus-Based Materials for the Language Classroom In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. xxviii, 682 p.). London: Routledge.

Strevens, P. (1973). Technical, Technological, and Scientific English (TTSE)1. *ELT Journal, XXVII* (3), 223-234. doi:10.1093/elt/XXVII.3.223

Stubbs, M. (1993). British Traditions in Text Analysis-From Firth to Sinclair. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: in honour of John Sinclair*. Philadelphia J. Benjamins Pub. Co.

Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford; Malden, MA: Blackwell Publishers.

Stubbs, M. (2007). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Language and computers no. 62* (pp. 385 p.). Amsterdam: Rodopi.

Stubbs, M. (2007). On texts, corpora and model of language. In M. Hoey, M. Mahlberg, M. Stubbs, & W. Teubert (Eds.), *Text, discourse and corpora theory and analysis*. London; New York: Continuum.

Stubbs, M. (2007). Quantitative data on multi-word sequences in English: the case of the word *world*. In M. Hoey, M. Mahlberg, M. Stubbs, & W. Teubert (Eds.), *Text, discourse and corpora: theory and analysis* (pp. x, 253 p.). London: Continuum.

Svartvik, J. (1990). *The London-Lund corpus of spoken English: description and research*. Lund, Sweden: Lund U.P.

Svensson, H. M. (2008). A very complex criterion of fixedness: Non-compositionality. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective* (pp.

xxviii, 422 p.). Amsterdam: John Benjamins.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge [England]; New York: Cambridge University Press.

Swales, J. (2004). *Research genres: explorations and applications*. Cambridge: Cambridge University Press.

Teubert, W., & Cermáková, A. (2007). *Corpus linguistics: a short introduction*. London: Continuum.

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

Tomlinson, B. (2011). *Materials development in language teaching* (2nd ed.). Cambridge: Cambridge University Press.

Tomlinson, B., & Masuhara, H. (2011). *Research for Materials Development in Language Learning: Evidence For Best Practice* (B. Tomlinson & H. Masuhara Eds.): Continuum

Tronenko, N. G. (2001). *Regularities in the behaviour of Russian phrasal idioms* (Doctoral dissertation, University of Oxford).

Ur, P. (1996). *A Course in language teaching: practice and theory*. Cambridge: Cambridge University Press.

Vyatkina, N., & Boulton, A. (2017). Corpora in language teaching and learning. *Language learning and technology*, *21*(3), 66-89. DOI:10.1017/S0261444817000167

Wang, J., Liang, S.-l., & Ge, G.-c. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes, 27*(4), 442-458. DOI:10.1016/j.esp.2008.05.003

West, M. (1953). A general service list of English words. London: Longman, Green and Co.

Weinert, R. (1995). The Role of Formulaic Language in Second Language Acquisition: A Review. *Applied Linguistics, 16*(2), 180-205. https://doi.org/10.1093/applin/16.2.180

Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Edward Arnold.

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

Yüzlü, M. Y., & Atay, D. (2020). The effectiveness of L1 use in promoting oral production in L2 across different proficiency levels. *Journal of Language and Linguistic Studies*, *16*(2), 864-882. DOI:10.17263/jlls.759330

Xiao, R. (2015). Collocation. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge, UK: Cambridge University Press.

Appendix (1). English Courses Descriptions provided for engineering students

01355204* การฟัง-การพูดภาษาอังกฤษเบื้องต้น 3(3-0-6)
(Fundamental English Listening-Speaking)
พื้นฐาน : 01355113 หรือกำหนดจากผลสอบเข้ามหาวิทยาลัยเฉพาะคะแนนวิชาภาษาอังกฤษ
การฝึกทักษะฟัง-พูด โดยใช้กิจกรรมหลากหลายรูปแบบในหัวข้อต่างๆ โดยมุ่งเน้นให้ผู้เรียนฝึกทักษะภาษาอังกฤษ
Practicing listening and speaking skills through various activities in a variety of topics with an emphasis on helping students practice their English.

01355205** การอ่านภาษาอังกฤษด้านสื่อสารมวลชน 3(3-0-6)
(Reading for Mass Communication in English)
พื้นฐาน : 01355113 หรือกำหนดจากผลสอบเข้ามหาวิทยาลัยเฉพาะคะแนนวิชาภาษาอังกฤษ
การอ่านสิ่งตีพิมพ์ภาษาอังกฤษด้านสื่อสารมวลชน ข่าว โฆษณา บทความ หนังสือพิมพ์และนิตยสาร ข่าวจากโทรพิมพ์ และการพิสูจน์อักษร โดยเน้นวิธีอ่านเพื่อจับใจความสำคัญ สำนวนและลีลาการเสนอข่าวและบทความ
Reading mass media materials such as news, advertisements, newspaper and magazine articles, teletype news and proofreading. Emphasis is given to reading techniques to comprehend main ideas, vocabulary, idiomatic expressions and news and feature writing styles.

01355206** อังกฤษวิชาการ 3(3-0-6)
(Technical English)
พื้นฐาน : 01355113 หรือกำหนดจากผลสอบเข้ามหาวิทยาลัยเฉพาะคะแนนวิชาภาษาอังกฤษ
ฝึกฝนทักษะการอ่านเพื่อให้นิสิตได้คุ้นเคยกับศัพท์ในตำราเรียน วารสาร และสิ่งตีพิมพ์อื่น ๆ เกี่ยวกับวิชาการต่าง ๆ
Practicing reading skill in order to familiarize the students with technical terms found in technical textbooks of various fields, periodicals and other printed matter.

01355207** การเขียนโต้ตอบภาษาอังกฤษ 3(3-0-6)
(English Correspondence)
พื้นฐาน : 01355113 หรือกำหนดจากผลสอบเข้ามหาวิทยาลัยเฉพาะคะแนนวิชาภาษาอังกฤษ
การเขียนจดหมายประเภทต่าง ๆ เน้นในเรื่องแบบฟอร์ม ศัพท์ สำนวน ศิลปะการเขียนจดหมายสมัครงานและวิธีการทำประวัติส่วนตัว
Writing of various types of letters with emphasis on form, vocabulary and idiomatic and conventional expressions, including how to write letters of application and prepare resumes.

---

* วิชาเปิดใหม่
** วิชาปรับปรุง

Appendix (2). Sources of Engineering Research Articles (ERAs) analysed in the study

No.	Engineering	Journal	Year
	Civil Engineering		
1	Limestone filler cement in low w/c concrete: A rational use of energy	Cement and Concrete Research	2003
2	Properties of volcanic pumice based cement and lightweight concrete	Cement and Concrete Research	2004
3	Cement paste colouring in concretes	Cement and Concrete Research	2004
4	Rheological behaviour of fresh cement pastes formulated from a Self Compacting Concrete (SCC)	Cement and Concrete Research	2006
5	A new model for the estimation of compressive strength of Portland cement concrete	Cement and Concrete Research	2006
6	Performance of volcanic ash and pumice based blended cement concrete in mixed sulfate environment	Cement and Concrete Research	2006
7	The influence of potassium-sodium ratio in cement on concrete expansion due to alkali-aggregate reaction	Cement and Concrete Research	2008
8	Investigation on key properties controlling early-age stress development of blended cement concrete	Cement and Concrete Research	2008
9	Cement content determination through selective stain in hardened concrete	Cement and Concrete Research	2009
10	Corrosion initiation of reinforced concretes based on Portland or GGBS cements: Chloride contents and electrochemical characterizations versus time	Cement and Concrete Research	2012
11	Preferential adsorption of polycarboxylate superplasticizers on cement and silica fume in ultra-high performance concrete (UHPC)	Cement and Concrete Research	2012
12	An experimental study on cracking evolution in concrete and cement mortar by the b-value analysis of acoustic emission technique	Cement and Concrete Research	2013
13	Eco-friendly concretes with reduced water and cement contents – Mix design principles and laboratory tests	Cement and Concrete Research	2013
14	A new US procedure to determine setting period of cement pastes, mortars, and concretes	Cement and Concrete Research	2013
15	Mix design and properties assessment of Ultra-High Performance Fibre Reinforced Concrete (UHPFRC)	Cement and Concrete Research	2014
16	MgO expansive cement and concrete in China: Past, present and future	Cement and Concrete Research	2014
17	Rice husk ash (RHA) effectiveness in cement and concrete as a function of reactive silica and fineness	Cement and Concrete Research	2014
18	Chloride concentration in the pore solution of Portland cement paste and Portland cement concrete	Cement and Concrete Research	2014
19	Effect of superabsorbent polymers (SAPs) on rheological properties of fresh cement-based mortars – Development of yield stress and plastic viscosity over time	Cement and Concrete Research	2015
20	Mathematical modeling of CO2 uptake by concrete during accelerated carbonation curing	Cement and Concrete Research	2015
21	Reliable specific surface area measurements on anhydrous cements	Cement and Concrete Research	2015
22	Impact of chloride-rich environments on cement paste mineralogy	Cement and Concrete Research	2015
23	Microwave processing of cement and concrete materials - towards an industrial reality?	Cement and Concrete Research	2015
24	The stability of bound chlorides in cement paste with sulfate attack	Cement and Concrete Research	2015
25	Influence of the Ca/Si ratio of the C-S-H phase on the interaction with sulfate ions and its impact on the ettringite crystallization pressure	Cement and Concrete Research	2015

No.	Engineering	Journal	Year
	Computer and Electrical Engineering		
26	Message latency in hypercubic computer networks with the bursty traffic pattern	Computers and Electrical Engineering	2004
27	A hybrid intrusion detection system design for computer network security	Computers and Electrical Engineering	2009
28	Survey and analysis on Security Requirements Engineering	Computers and Electrical Engineering	2012
29	Partial bitstream protection for low-cost FPGAs with physical unclonable function, obfuscation, and dynamic partial self reconfiguration	Computers and Electrical Engineering	2013
30	An automatic computer-aided diagnosis system for liver tumours on computed tomography images	Computers and Electrical Engineering	2013
31	Adaptive workload driven dynamic power management for high performance computing clusters	Computers and Electrical Engineering	2013
32	Design of highly realistic virtual environment for excavator simulator	Computers and Electrical Engineering	2013
33	Delivery reliability of computer networks for data transmission within the permitted packet error rate and latency	Computers and Electrical Engineering	2013
34	Behavioral-model-based freehand tracking in a Selection-Move-Release system	Computers and Electrical Engineering	2014
35	Fast image processing for optical metrology utilizing heterogeneous computer architectures	Computers and Electrical Engineering	2014
36	Quantum computing and communications – Introduction and challenges	Computers and Electrical Engineering	2014
37	Evaluation of commercial brain–computer interfaces in real and virtual world environment: A pilot study	Computers and Electrical Engineering	2014
38	Secure embedded system hardware design – A flexible security and trust enhanced approach	Computers and Electrical Engineering	2014
39	Graphics Processing Units and Open Computing Language for parallel computing	Computers and Electrical Engineering	2014
40	Simulation of the smart grid communications: Challenges, techniques, and future trends	Computers and Electrical Engineering	2014
41	Evaluation of performance and architectural efficiency of FPGAs and GPUs in the 40 and 28 nm generations for algorithms in 3D ultrasound computer tomography	Computers and Electrical Engineering	2014
42	Contextual modeling for logical labeling of PDF documents	Computers and Electrical Engineering	2014
43	A survey on energy-efficient methodologies and architectures of network-on-chip	Computers and Electrical Engineering	2014
44	A novel approach to fault diagnosis for time-delay systems	Computers and Electrical Engineering	2014
45	Smart wireless sensor networks for online faults diagnosis in induction machine	Computers and Electrical Engineering	2014

No.	Engineering	Journal	Year
46	Using Software Defined Networking to manage and control IEC 61850-based systems	Computers and Electrical Engineering	2014
47	A cloud based and Android supported scalable home automation system	Computers and Electrical Engineering	2014
48	A novel algorithm using affine-invariant features for pose-variant face recognition	Computers and Electrical Engineering	2015
49	Application of van der Pol-Duffing oscillator in weak signal detection	Computers and Electrical Engineering	2015
50	A new method for removal of powerline interference in ECG and EEG recordings	Computers and Electrical Engineering	2015
	Electrical Power Engineering		
51	A novel hysteresis current control for active power filter with constant frequency	Electric Power Systems Research	2004
52	A standard method for specifying the response of hydroelectric plant in frequency-control mode	Electric Power Systems Research	2004
53	Evaluating operational risk in a power system with a large amount of wind power	Electric Power Systems Research	2009
54	Extending the perturbation technique to the modal representation of nonlinear systems	Electric Power Systems Research	2009
55	A novel frequency tracking method based on complex adaptive linear neural network state vector in power systems	Electric Power Systems Research	2009
56	Optimal spares availability strategy for power transformer components	Electric Power Systems Research	2010
57	Assessment of techno-economic contribution of FACTS devices to power system operation	Electric Power Systems Research	2010
58	Fault section estimation in electric power systems using an optimization immune algorithm	Electric Power Systems Research	2010
59	Stability analysis of five-phase induction motor drives with variable third harmonic injection	Electric Power Systems Research	2010
60	A novel domestic electric water heater model for a multi-objective demand side management program	Electric Power Systems Research	2010
61	A unified approach for the solution of power flows in electric power systems including wind farms	Electric Power Systems Research	2011
62	Identification of excitation systems with the generator online	Electric Power Systems Research	2012
63	Analysis of sequences of events for the characterisation of faults in power systems	Electric Power Systems Research	2012
64	Implementing agent-based emissions trading for controlling Virtual Power Plant emissions	Electric Power Systems Research	2013
65	A review of voltage dip mitigation techniques with distributed generation in electricity networks	Electric Power Systems Research	2013
66	A methodology for real time analysis of parallelism of distribution networks	Electric Power Systems Research	2013
67	Toward smart distribution management by integrating advanced metering infrastructure	Electric Power Systems Research	2013
68	Implementation of fractional slot concentrated winding technique to large salient-pole synchronous generators & development with permanent magnets	Electric Power Systems Research	2013
69	Large-scale control of domestic refrigerators for demand peak reduction in distribution systems	Electric Power Systems Research	2013
70	Simulative and experimental investigation of transfer function of inter-turn faults in transformer windings	Electric Power Systems Research	2014
71	Short-term scheduling of thermal power systems using hybrid gradient based modified teaching-learning optimizer with black hole algorithm	Electric Power Systems Research	2014
72	Dynamic Lightning Protection of Smart Grid distribution system	Electric Power Systems Research	2014
73	Energy recovery effectiveness in trolleybus transport	Electric Power Systems Research	2014
74	Enel Distribuzione projects for renewable energy sources integration in distribution grid	Electric Power Systems Research	2015

No.	Engineering	Journal	Year
75	EV fast charging stations and energy storage technologies: A real implementation in the smart micro grid paradigm	Electric Power Systems Research	2015
	Mechanical Engineering		
76	Induction heating apparatus for high temperature testing of thermo-mechanical properties	Applied Thermal Engineering	2009
77	Numerical solutions for functionally graded solids under thermal and mechanical loads using a high-order control volume finite element method	Applied Thermal Engineering	2011
78	PEEK film heat transfer surfaces for multi-effect distillation: A mechanical investigation	Applied Thermal Engineering	2012
79	Adsorbent coatings for heat pumping applications: Verification of hydrothermal and mechanical stabilities	Applied Thermal Engineering	2013
80	Electrical-fluid dynamic performance of mechanical draft water cooling towers	Applied Thermal Engineering	2013
81	Experimental study on the performance of a mechanical cooling tower fitted with different types of water distribution systems and drift eliminators	Applied Thermal Engineering	2013
82	Effect of mechanical vibration on flow and heat transfer characteristics in rectangular microgrooves	Applied Thermal Engineering	2013
83	Extended infrared thermography applied to orthogonal cutting: Mechanical and thermal aspects	Applied Thermal Engineering	2014
84	Thermal failure of rubber bushing of a Positive Displacement Motor: A study based on thermo-mechanical coupling	Applied Thermal Engineering	2014
85	Investigation of effect of process parameters on multilayer builds by direct metal deposition	Applied Thermal Engineering	2014
86	Characterization and modeling of a scroll expander with air and ammonia as working fluid	Applied Thermal Engineering	2014
87	Thermally induced mechanical response of energy piles in axially loaded pile groups	Applied Thermal Engineering	2014
88	Effect of various leg geometries on thermo-mechanical and power generation performance of thermoelectric devices	Applied Thermal Engineering	2014
89	Estimation of thermoelectric and mechanical performances of segmented thermoelectric generators under optimal operating conditions	Applied Thermal Engineering	2014
90	Morphology, physical, thermal and mechanical properties of the constitutive materials of diesel particulate filters	Applied Thermal Engineering	2014
91	Exergetic, economic, and environmental evaluations and multi-objective optimization of a combined molten carbonate fuel cell-gas turbine system	Applied Thermal Engineering	2015
92	Heat transfer characteristics of a circular tube bank fin heat exchanger with fins punched curve rectangular vortex generators in the wake regions of the tubes	Applied Thermal Engineering	2015
93	Integrated system of mechanical refrigeration and thermosyphon for free cooling of data centers	Applied Thermal Engineering	2015
94	Reduced scale thermal characterization of automotive disc brake	Applied Thermal Engineering	2015
95	Discussion of the internal heat exchanger's effect on the Organic Rankine Cycle	Applied Thermal Engineering	2015
96	Heat-transfer improvements in an axial-flux permanent-magnet synchronous machine	Applied Thermal Engineering	2015
97	Experimental performance of a rotating two-phase reaction turbine	Applied Thermal Engineering	2015
98	Thermodynamic optimisation and analysis of four Kalina cycle layouts for high temperature applications	Applied Thermal Engineering	2015
99	Performance of a rotating two-phase turbine for combined power generation and desalination	Applied Thermal Engineering	2015
100	Hybrid ANN-PLS approach to scroll compressor thermodynamic performance prediction	Applied Thermal Engineering	2015

Appendix (3). All 200 of the 3-word, 4-word lexical bundles from the top and bottom occur at least two times in two different texts in the ERAs corpus (per 400, 000 words). They were normalised.

No	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information: Structural Type	Discourse Function
1	shown in fig	400	1000	82	VP-based	Referential C3) Text deixis
2	in order to	358	895	74	PP-based	Discourse organiser B) Topic elaboration/clarification
3	due to the	276	690	78	Others	Discourse organiser B) Topic elaboration/clarification
4	based on the	229	572.5	73	VP-based	Discourse organiser B) Topic elaboration/clarification
5	the number of	202	505	54	NP-based	Referential B1) Quantifying specification
6	as shown in	192	480	58	PP-based	Referential C3) Text deixis
7	it can be	171	427.5	62	VP-based	Stance B) Attitudinal/modality stance B2) Ability
8	as well as	170	425	56	Others	Discourse organiser B) Topic elaboration/clarification
9	the use of	161	402.5	59	NP-based	Referential B3) Intangible framing attributes
10	of the system	144	360	40	PP-based	Referential C1) Place reference
11	the effect of	139	347.5	42	NP-based	Referential B3) Intangible framing attributes
12	according to the	136	340	60	PP-based	Discourse organiser A) Topic introduction/focus
13	in terms of	129	322.5	52	PP-based	Referential B3) Intangible framing attributes
14	is shown in	126	315	58	VP-based	Referential C3) Text deixis

15	in this study	119	297.5	42	PP-based	Discourse organiser A) Topic introduction/focus
16	in this paper	118	295	55	PP-based	Discourse organiser A) Topic introduction/focus
17	can be used	113	282.5	56	VP-based	Stance B) Attitudinal/modality stance B2) Ability
18	are shown in	112	280	54	VP-based	Referential C3) Text deixis
19	is used to	104	260	44	VP-based	Discourse organiser B) Topic elaboration/clarification
20	as a result	102	255	43	PP-based	Discourse organiser A) Topic introduction/focus
21	one of the	100	250	55	NP-based	Referential A) Identification/Focus
22	the presence of	100	250	39	NP-based	Referential B3) Intangible framing attributes
23	can be seen	99	247.5	46	VP-based	Stance B) Attitudinal/modality stance B2) Ability
24	the performance of	98	245	43	NP-based	Referential B3) Intangible framing attributes
25	used in the	97	242.5	42	VP-based	Discourse organiser B) Topic elaboration/clarification

No	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information: Structural Type	Discourse Function
26	of the proposed	95	237.5	34	PP-based	Referential C1) Place reference
27	on the other	95	237.5	48	PP-based	Discourse organiser B) Topic elaboration/clarification
28	with respect to	94	235	42	PP-based	Discourse organiser B) Topic elaboration/clarification
29	the development of	93	232.5	33	NP-based	Referential B3) Intangible framing attributes
30	compared to the	89	222.5	36	VP-based	Discourse organiser B) Topic elaboration/clarification
31	the value of	87	217.5	36	NP-based	Referential B1) Quantifying specification
32	the other hand	86	215	47	NP-based	Discourse organiser B) Topic elaboration/clarification
33	the amount of	86	215	35	NP-based	Referential B1) Quantifying specification
34	in the case	83	207.5	32	PP-based	Discourse organiser A) Topic introduction/focus
35	is based on	82	205	43	VP-based	Discourse organiser B) Topic elaboration/clarification
36	in case of	82	205	19	PP-based	Referential B3) Intangible framing attributes
37	at the same	81	202.5	44	PP-based	Referential B3) Intangible framing attributes
38	efficiency of the	81	202.5	23	NP-based	Referential B3) Intangible framing attributes
39	the heat transfer	81	202.5	14	Others	Other
40	in this case	75	187.5	41	PP-based	Discourse organiser A) Topic introduction/focus
41	values of the	74	185	37	NP-based	Referential B3) Intangible framing attributes
42	to determine the	73	182.5	49	Clausal	Discourse organiser A) Topic introduction/focus

43	a function of	73	182.5	32	NP-based	Referential B3) Intangible framing attributes
44	the case of	72	180	34	NP-based	Referential B3) Intangible framing attributes
45	shown in table	71	177.5	44	VP-based	Referential C3) Text deixis
46	performance of the	71	177.5	35	NP-based	Referential B3) Intangible framing attributes
47	be used to	69	172.5	40	VP-based	Discourse organiser B) Topic elaboration/clarification
48	part of the	69	172.5	38	NP-based	Referential B1) Quantifying specification
49	value of the	69	172.5	36	NP-based	Referential B1) Quantifying specification
50	related to the	67	167.5	42	VP-based	Discourse organiser B) Topic elaboration/clarification

NO	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information : Structural Type	Discourse Function
51	as shown in fig	141	352.5	44	PP-based	Referential C3) Text deixis
52	is shown in fig	114	285	55	VP-based	Referential C3) Text deixis
53	on the other hand	86	215	47	PP-based	Discourse organiser B) Topic elaboration/clarification
54	are shown in fig	72	180	35	VP-based	Referential C3) Text deixis
55	in the case of	61	152.5	31	PP-based	Referential B3) Intangible framing attributes
56	as well as the	58	145	29	Others	Discourse organiser B) Topic elaboration/clarification
57	it can be seen	57	142.5	25	VP-based	Stance B) Attitudinal/modality stance B2) Ability
58	at the same time*	56	140	34	PP-based	Referential B3) Intangible framing attributes/Discourse organiser A) Topic introduction/focus
59	as a function of	54	135	23	PP-based	Referential B3) Intangible framing attributesgible framing attributes
60	as a result of	52	130	25	PP-based	Discourse organiser A) Topic introduction/focus
61	with respect to the	47	117.5	32	PP-based	Discourse organiser B) Topic elaboration/clarification
62	the effect of the	45	112.5	20	NP-based	Referential B3) Intangible framing attributes
63	can be used to	44	110	29	VP-based	Stance B) Attitudinal/modality stance B2) Ability
64	it is possible to	43	107.5	27	VP-based	Stance B) Attitudinal/modality stance B2) Ability
65	the performance of the	38	95	23	NP-based	Referential B3) Intangible framing attributes

66	can be found in	38	95	22	VP-based	Stance B) Attitudinal/modality stance B2) Ability
67	in terms of the	37	92.5	16	PP-based	Referential B3) Intangible framing attributes
68	is based on the	36	90	23	VP-based	Discourse organiser B) Topic elaboration/clarification
69	the temperature of the	34	85	13	NP-based	Referential B1) Quantifying specification
70	a result of the	33	82.5	19	NP-based	Referential B3) Intangible framing attributes
71	in case of the	33	82.5	7	PP-based	Referential B3) Intangible framing attributes
72	in this study the	31	77.5	20	PP-based	Discourse organiser A) Topic introduction/focus
73	can be seen that	31	77.5	17	VP-based	Stance B) Attitudinal/modality stance B2) Ability
74	in the presence of	31	77.5	17	PP-based	Referential B3) Intangible framing attributes
75	the size of the	30	75	22	NP-based	Referential B1) Quantifying specification

No	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information: Structural Type	Discourse Function
76	it is necessary to	30	75	20	VP-based	Stance B) Attitudinal/modality stance B1) Obligation/ directive
77	it should be noted*	29	72.5	21	VP-based	Stance B. Attitudinal/Modality stance/Discourse organiser B) Topic elaboration/clarification B. 1) Obligation/directive stance: Impersonal/Discourse organiser B) Topic elaboration/clarification
78	the end of the	29	72.5	16	NP-based	Referential C1) Place reference/C2) Time reference
79	the beginning of the	29	72.5	15	NP-based	Referential C1) Place reference/C2) Time reference
80	on the basis of	29	72.5	13	PP-based	Referential B2) Intangible framing attributes
81	a function of the	28	70	19	NP-based	Referential B2) Intangible framing attributes
82	it is important to	28	70	19	VP-based	Stance B) Attitudinal/modality stance B1) Obligation/ directive
83	in the range of	28	70	15	PP-based	Referential B3) Intangible framing attributes
84	are shown in table	27	67.5	21	VP-based	Referential C3) Text deixis
85	in this case the	27	67.5	18	PP-based	Discourse organiser A) Topic introduction/focus
86	as a result the	27	67.5	17	PP-based	Referential B3) Intangible framing attributes
87	can be seen from	27	67.5	13	VP-based	Stance B) Attitudinal/modality stance B2) Ability
88	should be noted that	26	65	20	VP-based	Discourse organiser B) Topic elaboration/clarification
89	in this paper the	26	65	21	PP-based	Discourse organiser A) Topic introduction/focus

90	is due to the	26	65	19	VP-based	Discourse organiser B) Topic elaboration/clarification
91	the other hand the	26	65	19	NP-based	Discourse organiser B) Topic elaboration/clarification
92	at the end of*	24	60	17	PP-based	Referential C1) Place reference/C2) Time reference
93	in the form of	24	60	18	PP-based	Referential B1) Quantifying specification
94	be attributed to the	24	60	11	VP-based	Discourse organiser B) Topic elaboration/clarification
95	can be concluded that	23	57.5	16	VP-based	Stance B) Attitudinal/modality stance B2) Ability
96	the evolution of the	24	60	10	NP-based	Referential B3) Intangible framing attributes
97	to the fact that	24	60	16	PP-based	Stance A) Epistemic
98	are listed in table	24	60	16	VP-based	Referential C3) Text deixis
99	it can be concluded	23	57.5	16	VP-based	Stance B) Attitudinal/modality stance B2) Ability/Discourse organiser
100	an increase in the	24	60	13	NP-based	Referential B3) Intangible framing attributes

No	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information: Structural Type	Discourse Function
101	flow analysis and	2	5	2	NP-based	Others
102	flow and heat	6	15	2	NP-based	Others
103	flow can be	2	5	2	VP-based	Stance B) Attitudinal/modality stance B2) Ability
104	flow chart is	2	5	2	VP-based	Others
105	flow chart of	2	5	2	NP-based	Referential B2) Tangible framing attributes
106	flow heat exchanger	2	5	2	NP-based	Others
107	flow into the	3	7.5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
108	flow of air	2	5	2	NP-based	Referential B3) Intangible framing attributes
109	flow per unit	2	5	2	NP-based	Referential B1) Quantifying specification
110	flow rate can	2	5	2	VP-based	Stance B) Attitudinal/modality stance B2) Ability
111	flow rates of	2	5	2	NP-based	Referential B1) Quantifying specification
112	flow ratio are	2	5	2	VP-based	Referential B1) Quantifying specification
113	flow reaction turbine	2	5	2	Others	Others
114	flow resistance in	2	5	2	NP-based	Referential C1) Place reference
115	flow simple reaction	4	10	2	NP-based	Others
116	flow value of	2	5	2	NP-based	Referential B1) Quantifying specification
117	flows into the	2	5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
118	flows through the	2	5	2	NP-based	Discourse organiser B) Topic elaboration/clarification

119	influence of thermal	3	7.5	2	NP-based	Referential B3) Intangible framing attributes
120	influences on the	2	5	2	NP-based	Referential B3) Intangible framing attributes
121	fluid and angular	2	5	2	NP-based	Others
122	fluid as it	2	5	2	NP-based	Others
123	fluid exiting the	3	7.5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
124	fluid leaves the	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
125	fluid leaving the	2	5	2	NP-based	Discourse organiser B) Topic elaboration/clarification

No	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information: Structural Type	Discourse Function
126	flux distribution in	2	5	2	NP-based	Referential B3) Intangible framing attributes
127	flux in the	3	7.5	2	NP-based	Referential C1) Place reference
128	flux per unit	3	7.5	2	NP-based	Referential B1) Quantifying specification
129	fly ash and	8	20	2	NP-based	Others
130	fly ash mortars	2	5	2	NP-based	Others
131	fly ash on	2	5	2	NP-based	Referential C1) Place reference
132	briefly explained in	2	5	2	VP-based	Referential C1) Place reference
133	briefly summarised below	2	5	2	VP-based	Referential C1) Place reference
134	xrd analysis of	2	5	2	NP-based	Referential B2) Tangible framing attributes
135	yield stress are	4	10	2	VP-based	Others
136	yield stress in	6	15	2	NP-based	Referential C1) Place reference
137	yield stress is	2	5	2	VP-based	Others
138	yield stress than	2	5	2	NP-based	Others
139	yield stress to	2	5	2	NP-based	Others
140	yield stress values	12	30	2	NP-based	Others
141	yields the following	3	7.5	2	VP-based	Others
142	young concretes with	4	10	2	NP-based	Others
143	zero at the	2	5	2	NP-based	Referential C1) Place reference
144	zero in the	2	5	2	NP-based	Referential C1) Place reference
145	zone and the	3	7.5	2	NP-based	Others

146	zone between the	2	5	2	NP-based	Referential C1) Place reference
147	zone of the	2	5	2	NP-based	Referential C1) Place reference
148	zone to the	2	5	2	NP-based	Referential C1) Place reference
149	reflected in the	2	5	2	VP-based	Others
150	flexibility in the	2	5	2	NP-based	Referential C1) Place reference

No	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information: Structural Type	Discourse Function
151	efficiency as compared to	3	7.5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
152	efficiency due to the	2	5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
153	efficiency of the expander	10	25	2	NP-based	Referential B2) Tangible framing attributes
154	efficiency of the plant	5	12.5	2	NP-based	Referential B2) Tangible framing attributes
155	efficiency of the turbine	8	20	2	NP-based	Referential B2) Tangible framing attributes
156	efficiency was found to	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
157	efficient and reliable communication	2	5	2	NP-based	Others
158	difficult to manage when	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
159	figure shows that the	2	5	2	VP-based	Discourse organiser A) Topic introduction/focus
160	profile along the distribution	2	5	2	NP-based	Referential C1) Place reference
161	filter paper and the	2	5	2	NP-based	Others
162	finding is that the	3	7.5	2	VP-based	Discourse organiser A) Topic introduction/focus
163	findings can be summarised	2	5	2	VP-based	Stance B) Attitudinal/modality stance B2) Ability
164	defined according to the	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
165	defined in terms of	2	5	2	VP-based	Referential B3) Intangible framing attributes
166	defined three types of	2	5	2	VP-based	Referential B3) Intangible framing attributes
167	finite element model of	2	5	2	NP-based	Referential B3) Intangible framing attributes

168	first and the second	2	5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
169	first is that the	2	5	2	VP-based	Discourse organiser A) Topic introduction/focus
170	first law of thermodynamics	2	5	2	NP-based	Referential B3) Intangible framing attributes
171	first to use the	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
172	flash evaporates in the	2	5	2	VP-based	Referential C1) Place reference
173	flash evaporation from a	2	5	2	NP-based	Referential C1) Place reference
174	flashing of superheated liquid	2	5	2	NP-based	Referential B2) Tangible framing attributes
175	flashing of the feed	2	5	2	NP-based	Referential B2) Tangible framing attributes
No	Bundle	Frequency (per 400,000 words)	ERAs corpus dataNormalised (per million words)	Dispersion Range (100 texts in total)	Bundle information: Structural Type	Discourse Function
176	flashing tank is between	2	5	2	VP-based	Referential C1) Place reference
177	flow and heat transfer	2	5	2	NP-based	Others
178	flow chart of the	2	5	2	NP-based	Referential B2) Tangible framing attributes
179	flow rate and temperature	2	5	2	NP-based	Others
180	flow rate and the	2	5	2	NP-based	Others
181	flow rate can be	2	5	2	VP-based	Stance B) Attitudinal/modality stance B2) Ability
182	flow rate increases due	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
183	flow rate of air	2	5	2	NP-based	Referential B1) Quantifying specification

184	flow rate of working	2	5	2	NP-based	Referential B1) Quantifying specification
185	flow rate was measured	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
186	flow reaction turbine rotor	2	5	2	NP-based	Others
187	flow resistance in the	2	5	2	NP-based	Referential C1) Place reference
188	flow simple reaction turbines	2	5	2	NP-based	Others
189	flows to the condenser	2	5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
190	influence of the temperature	2	5	2	NP-based	Referential B3) Intangible framing attributes
191	fluid and angular speed	2	5	2	NP-based	Others
192	fluid as it leaves	2	5	2	VP-based	Others
193	fluid enters the condenser	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
194	fluid exiting the turbine	3	7.5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
195	fluid leaves the turbine	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
196	fluid leaving the turbine	13	32.5	2	NP-based	Discourse organiser B) Topic elaboration/clarification
197	yield stress and a	2	5	2	NP-based	Others
198	yield stress and plastic	7	17.5	2	NP-based	Others
199	specific surface area increases	2	5	2	VP-based	Discourse organiser B) Topic elaboration/clarification
200	efficacy of the proposed	2	5	2	NP-based	Referential B3) Intangible framing attributes

Appendix (4). Inter-rater reliability: structural and functional categorisation of the 200 bundles from the Top 50 and Bottom 50 of 3-word and 4-word bundles

No.	Lexical bundles	Frequency	Structural Type	NP	PP	VP	CF	Other	Rater 1	Rater 2	Coding
1	shown in fig	400	VP-based			1			6	6	1
2	in order to	358	PP-based		1				4	4	1
3	due to the	276	PP-based		1				4	4	1
4	based on the	229	VP-based			1			6	6	1
5	the number of	202	NP-based	1					1	1	1
6	as shown in	192	Clausal fragment				1		9	9	1
7	it can be	171	VP-based			1			10	10	1
8	as well as	170	Other					1	14	14	1
9	the use of	161	NP-based	1					1	1	1
10	of the system	144	PP-based		1				4	4	1
11	the effect of	139	NP-based	1					1	1	1
12	according to the	136	PP-based		1				4	4	1
13	in terms of	129	PP-based		1				3	3	1
14	is shown in	126	VP-based			1			6	6	1
15	in this study	119	PP-based		1				4	4	1
16	in this paper	118	PP-based		1				4	4	1
17	can be used	113	VP-based			1			5	5	1
18	are shown in	112	VP-based			1			6	6	1
19	is used to	104	VP-based			1			6	5	0
20	as a result	102	Other					1	4	14	0
21	one of the	100	NP-based	1					1	1	1
22	the presence of	100	NP-based	1					1	1	1
23	can be seen	99	VP-based			1			5	5	1
24	the performance of	98	NP-based	1					1	1	1
25	used in the	97	VP-based			1			6	6	1
26	of the proposed	95	PP-based		1				4	4	1
27	on the other	95	PP-based		1				4	4	1
28	with respect to	94	PP-based		1				4	4	1
29	the development of	93	NP-based	1					1	1	1
30	compared to the	89	VP-based			1			6	6	1
31	the value of	87	NP-based	1					1	1	1
32	the other hand	86	Other: phrase fragment					1	15	15	1
33	the amount of	86	NP-based	1					1	1	1
34	in the case	83	PP-based		1				4	4	1
35	is based on	82	VP-based			1			6	6	1
36	in case of	82	PP-based		1				3	3	1
37	at the same	81	PP-based		1				4	4	1
38	efficiency of the	81	NP-based	1					1	1	1
39	the heat transfer	81	Other					1	15	14	0
40	in this case	75	PP-based		1				4	4	1
41	values of the	74	NP-based	1					1	1	1
42	to determine the	73	Clausal fragment				1		13	13	1
43	a function of	73	NP-based	1					1	1	1
44	the case of	72	NP-based	1					1	1	1
45	shown in table	71	VP-based		1				6	6	1
46	performance of the	71	NP-based	1					1	1	1
47	be used to	69	VP-based		1				6	5	0
48	part of the	69	NP-based	1					1	1	1
49	value of the	69	NP-based	1					1	1	1
50	related to the	67	VP-based			1			6	6	1

No.	Lexical bundles	Frequency	Structural Type	NP	PP	VP	CF	Other	Rater 1	Rater 2	Coding
51	as shown in fig	141	Clausal fragment				1		9	9	1
52	is shown in fig	114	VP-based			1			6	6	1
53	on the other hand	86	PP-based		1				4	4	1
54	are shown in fig	72	VP-based			1			6	6	1
55	in the case of	61	PP-based		1				3	3	1
56	as well as the	58	Other: phrase fragment					1	15	15	1
57	it can be seen	57	VP-based			1			11	11	1
58	at the same time	56	PP-based		1				4	4	1
59	as a function of	54	PP-based		1				3	3	3
60	as a result of	52	PP-based		1				3	3	1
61	with respect to the	47	PP-based		1				4	4	1
62	the effect of the	45	NP-based	1					1	1	1
63	can be used to	44	VP-based			1			6	5	0
64	it is possible to	43	VP-based			1			10	10	1
65	the performance of the	38	NP-based	1					1	1	1
66	can be found in	38	VP-based			1			6	6	1
67	in terms of the	37	PP-based		1				3	3	1
68	is based on the	36	VP-based			1			6	6	1
69	the temperature of the	34	NP-based	1					1	1	1
70	a result of the	33	NP-based	1					1	1	1
71	in case of the	33	PP-based		1				3	3	1
72	in this study the	31	PP-based		1				4	4	1
73	can be seen that	31	Clausal fragment				1		12	12	1
74	in the presence of	31	PP-based		1				3	3	1
75	the size of the	30	NP-based	1					1	1	1
76	it is necessary to	30	VP-based			1			10	10	1
77	it should be noted	29	VP-based			1			11	11	1
78	the end of the	29	NP-based	1					1	1	1
79	the beginning of the	29	NP-based	1					1	1	1
80	on the basis of	29	PP-based		1				3	3	1
81	a function of the	28	NP-based	1					1	1	1
82	it is important to	28	VP-based			1			10	10	1
83	in the range of	28	PP-based		1				3	3	1
84	are shown in table	27	VP-based			1			6	6	1
85	in this case the	27	PP-based		1				4	4	1
86	as a result the	27	Other: phrase fragment					1	15	15	1
87	can be seen from	27	VP-based			1			6	6	1
88	should be noted that	26	Clausal fragment				1		12	12	1
89	in this paper the	26	PP-based		1				4	4	1
90	is due to the	26	VP-based			1			8	8	1
91	the other hand the	26	Other: phrase fragment					1	15	15	1
92	at the end of	24	PP-based		1				3	3	1
93	in the form of	24	PP-based		1				3	3	1
94	be attributed to the	24	VP-based			1			6	6	1
95	can be concluded that	23	Clausal fragment				1		12	12	1
96	the evolution of the	24	NP-based	1					1	1	1
97	to the fact that	24	PP-based		1				4	4	1
98	are listed in table	24	VP-based			1			6	6	1
99	it can be concluded	23	VP-based			1			11	11	1
100	an increase in the	24	NP-based	1					2	2	1
			Total								97%

No.	Lexical bundles	Rater 1	Rater 2	Coding
1	shown in fig	D. Deitics and locative	D. Deitics and locative	1
2	in order to	B. Topic Elaboration/ Clarification; cause and effect	B. Topic Elaboration/ Clarification; cause and effect	1
3	due to the	B. Topic Elaboration/ Clarification; cause and effect	B. Topic Elaboration/ Clarification; cause and effect	1
4	based on the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
5	the number of	B. 1) Quantifying specification	B. 1) Quantifying specification	1
6	as shown in	D. Deitics and locative	D. Deitics and locative	1
7	it can be	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
8	as well as	D. Discourse markers	D. Discourse markers	1
9	the use of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
10	of the system	D. Deitics and locative	D. Deitics and locative	0
11	the effect of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
12	according to the	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
13	in terms of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
14	is shown in	D. Deitics and locative	D. Deitics and locative	1
15	in this study	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
16	in this paper	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
17	can be used	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
18	are shown in	D. Deitics and locative	D. Deitics and locative	1
19	is used to	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
20	as a result	B. Topic Elaboration/ Clarification; cause and effect	B. Topic Elaboration/ Clarification; cause and effect	1
21	one of the	A. Identification/Focus	A. Identification/Focus	1
22	the presence of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
23	can be seen	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
24	the performance of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
25	used in the	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
26	of the proposed	D. Deitics and locative	D. Deitics and locative	0
27	on the other	D. Discourse markers	D. Discourse markers	1
28	with respect to	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
29	the development of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
30	compared to the	C. Contrast and comparison	C. Contrast and comparison	1
31	the value of	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
32	the other hand	D. Discourse markers	D. Discourse markers	1
33	the amount of	B. 1) Quantifying specification	B. 1) Quantifying specification	1
34	in the case	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
35	is based on	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
36	in case of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
37	at the same*	D. Discourse markers	D. Discourse markers	0
38	efficiency of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
39	the heat transfer	4. Other bundles in engineering	4. Other bundles in engineering	1
40	in this case	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
41	values of the	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
42	to determine the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
43	a function of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
44	the case of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
45	shown in table	D. Deitics and locative	D. Deitics and locative	1
46	performance of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
47	be used to	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
48	part of the	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
49	value of the	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
50	related to the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1

No.	Lexical bundles	Rater 1	Rater 2	Coding
51	as shown in fig	D. Deitics and locative	D. Deitics and locative	1
52	is shown in fig	D. Deitics and locative	D. Deitics and locative	1
53	on the other hand	D. Discourse markers	D. Discourse markers	1
54	are shown in fig	D. Deitics and locative	D. Deitics and locative	1
55	in the case of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
56	as well as the	D. Discourse markers	D. Discourse markers	1
57	it can be seen	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
58	at the same time	D. Discourse markers	D. Discourse markers	1
59	as a function of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
60	as a result of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
61	with respect to the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
62	the effect of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
63	can be used to	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
64	it is possible to	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
65	the performance of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
66	can be found in	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
67	in terms of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
68	is based on the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
69	the temperature of the	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
70	a result of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
71	in case of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
72	in this study the	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
73	can be seen that	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
74	in the presence of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
75	the size of the	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
76	it is necessary to	B. 1) Obligation/directive stance: Impersonal	B. 1) Obligation/directive stance: Impersonal	1
77	it should be noted	B. 1) Obligation/directive stance: Impersonal	B. 1) Obligation/directive stance: Impersonal	1
78	the end of the	D.2) Multi-functional reference	D.2) Multi-functional reference	1
79	the beginning of the	D.2) Multi-functional reference	D.2) Multi-functional reference	1
80	on the basis of	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
81	a function of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
82	it is important to	B. 1) Obligation/directive stance: Impersonal	B. 1) Obligation/directive stance: Impersonal	1
83	in the range of	B. 1) Quantifying specification	B. 1) Quantifying specification	1
84	are shown in table	D. Deitics and locative	D. Deitics and locative	1
85	in this case the	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
86	as a result the	B. Topic Elaboration/ Clarification: cause and effect	B. Topic Elaboration/ Clarification: cause and effect	1
87	can be seen from	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
88	should be noted that	B. 1) Obligation/directive stance: Impersonal	B. 1) Obligation/directive stance: Impersonal	1
89	in this paper the	A. Metadiscourse and textual reference	A. Metadiscourse and textual reference	1
90	is due to the	B. Topic Elaboration/ Clarification: cause and effect	B. Topic Elaboration/ Clarification: cause and effect	1
91	the other hand the	D. Discourse markers	D. Discourse markers	1
92	at the end of	D.2) Multi-functional reference	D.2) Multi-functional reference	1
93	in the form of	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
94	be attributed to the	B. Topic Elaboration/ Clarification: cause and effect	B. Topic Elaboration/ Clarification: cause and effect	1
95	can be concluded that	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
96	the evolution of the	B.3) Intangible framing attributes	B.3) Intangible framing attributes	1
97	to the fact that	B. Topic Elaboration/ Clarification: cause and effect	B. Topic Elaboration/ Clarification: cause and effect	1
98	are listed in table	D. Deitics and locative	D. Deictic and locative	1
99	it can be concluded	B.2) Ability: Impersonal	B.2) Ability: Impersonal	1
100	an increase in the	B.2) Tangible framing attributes	B.2) Tangible framing attributes	1
			Total	97%

Appendix (5). Participant information sheet

16 February 2016

PARTICIPANT INFORMATION SHEET

Title: Fixed sequences of words in Engineering Research Articles (ERAs): materials development for engineering students

Researcher: Mrs. Kamonchanok Sanmuang, k.sanmuang@lancaster.ac.uk

This study is being carried out as part of my Doctoral studies in the Department of Linguistics and English Language. It aims to assess the quality of the materials design to help improve engineering students' research article writing in English. Students' English language performance will be assessed with short tests at the start and at the end of the research. The students' perceptions of materials will be investigated. I have approached you because you are a member of the target population. I would be very grateful if you would agree to take part.

Participating in this series of sessions over 4 weeks involves 1) providing your educational background by filling in a one-page questionnaire, 2) performing a pre-test to assess your level of English 3) attending 4 training sessions (2.5 hours each) 4) performing a post-test to assess any improvement in your level of English, 5) filling out a questionnaire, and 6) participating in an oral interview regarding the effectiveness of the materials used. You will hopefully find participation interesting and it may be helpful to you in developing aspects of your writing in English.

You are free to withdraw from the study at any time and you do not have to give a reason. If you withdraw more than 2 weeks after the data is collected, the data will still be used for the project. At every stage, your name will remain confidential. Your real name will not be used but assigned a pseudonym. The data will be kept securely in a locked cupboard and electronic, and will be saved on a computer protected by password access, and the files will be encrypted. The data will be used for academic purposes only, primarily writing a PhD thesis, possibly also for conference presentations and academic publications such as journal articles.

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact myself or my supervisor, Dr. Karin Tusting, who can be contacted at k.tusting@lancaster.ac.uk or by phone on ++44(0)1524510825. You may also contact the Head of the Department, Prof. Greg Myers, at g.myers@lancaster.ac.uk or by phone on +44(0) 1524 592454.

Mrs. Kamonchanok Sanmuang **(k.sanmuang@lancaster.ac.uk)**

PhD student

Thank you for considering your participation in this project.

Lancaster University
Lancaster LA1 4YL
United Kingdom
Tel: +44 (0)1524 593045
Fax: +44 (0)1524 843085
http://www.ling.lancs.ac.uk

วันที่ 16 กุมภาพันธ์ 2559

เอกสารข้อมูลคำอธิบายสำหรับผู้เข้าร่วมเป็นอาสาสมัครในการวิจัย

ชื่องานวิจัย กลุ่มคำศัพท์ภาษาอังกฤษในงานวิจัยทางวิศวกรรมศาสตร์: การพัฒนาสื่อการเรียนเพื่อพัฒนาทักษะการเขียนงานวิจัยภาษาอังกฤษสำหรับนิสิต/นักศึกษาสาขาวิศวกรรมศาสตร์

นักวิจัย: นางกมลชนก แสนเมือง, อีเมล์ k.sanmuang@lancaster.ac.uk

งานวิจัยนี้ได้จัดทำขึ้นเพื่อเป็นส่วนหนึ่งของการศึกษาระดับปริญญาเอกของข้าพเจ้า ภาควิชาภาษาศาสตร์และภาษาอังกฤษ จุดประสงค์ของงานวิจัยเพื่อประเมินคุณภาพของการออกแบบสื่อการเรียน เพื่อปรับปรุงทักษะการเขียนงานวิจัยภาษาอังกฤษทางวิศวกรรมศาสตร์ของนิสิต/นักศึกษาสาขาวิศวกรรมศาสตร์ นอกจากนี้ระดับความสามารถทางภาษาอังกฤษของนิสิตจะได้รับการประเมินทั้งก่อนและหลังการเข้าร่วมโครงการ ตลอดจนความคิดเห็นของนิสิตที่มีต่อสื่อการเรียน นิสิตได้รับการติดต่อเพื่อเข้าร่วมในงานวิจัยนี้เนื่องจาก นิสิตคือกลุ่มเป้าหมายในงานวิจัยนี้ ซึ่งข้าพเจ้า เป็นบุคลากรของมหาวิทยาลัยเกษตรศาสตร์ และเป็นผู้ออกแบบสื่อการเรียนนี้ ข้าพเจ้ายินดีอย่างยิ่งถ้านิสิตสนใจเข้าร่วมในงานวิจัยนี้

นิสิตที่ตัดสินใจเข้าร่วมโครงการในครั้งนี้จะได้รับการอบรมที่มีระยะเวลารวมทั้งสิ้น 4 สัปดาห์ รวม 20 ชั่วโมง ซึ่งนิสิตจะเข้าร่วมกิจกรรมซึ่งประกอบไปด้วย 1) เข้ารับการทดสอบทางภาษาอังกฤษก่อนเข้าร่วมโครงการ 2) เข้าร่วมอบรมตลอดเวลาทั้ง 4 สัปดาห์ 3) เข้ารับการทดสอบทางภาษาอังกฤษหลังเข้าร่วมโครงการ 4) เข้าร่วมการสัมภาษณ์กลุ่มเพื่อให้ข้อมูลด้านประสิทธิภาพของสื่อการเรียนที่ใช้ในการอบรม ซึ่งนิสิตจะพบว่าโครงการนี้เป็นโครงการที่น่าสนใจ และเป็นประโยชน์ต่อทักษะทางการเขียนภาษาอังกฤษของนิสิต และหลังจากสิ้นสุดโครงการ นิสิตจะได้รับเงิน 500 บาท เป็นการแสดงความขอบคุณที่นิสิตได้สละเวลาเพื่อเข้าร่วมโครงการในครั้งนี้

นิสิตสามารถถอนตัวออกจากการวิจัยนี้ได้ตลอดเวลา และนิสิตไม่จำเป็นต้องชี้แจงเหตุผล ข้อมูลของนิสิตจะยังถูกนำมาใช้ประกอบในงานวิจัยถ้านิสิตถอนตัวสองอาทิตย์หลังจากเก็บข้อมูลทั้งหมดเรียบร้อยแล้ว ข้อมูลของนิสิตจะถูกเก็บเป็นความลับ นามแฝงของนิสิตจะนำมาใช้แทนชื่อจริง ข้อมูลของนิสิตจะถูกเก็บในตู้มีกุญแจล็อค และเก็บในคอมพิวเตอร์ที่มีรหัสผ่าน และชื่อไฟล์จะทำถูกทำให้เป็นรหัส ข้อมูลทั้งหมดจะถูกนำมาใช้เพื่อจุดประสงค์ทางการศึกษาเท่านั้น คือใช้ในการเขียนปริญญานิพนธ์ และใช้ในการนำเสนอผลงานในงานประชุมวิจัย และตีพิมพ์บทความทางวิชาการ เช่น วารสารทางวิชาการ

ถ้านิสิตมีข้อสงสัยหรือไม่พอใจกับในระหว่างเข้าร่วมในงานโครงการ นิสิตสามารถติดต่อข้าพเจ้าได้โดยตรง หรืออาจารย์ที่ปรึกษาโครงการวิจัยของข้าพเจ้า Dr. Karin Tusting อีเมล์ k.tusting@lancaster.ac.uk หรือโทรศัพท์ +44(0)1524510825 หรือสามารถติดต่อหัวหน้าภาควิชาภาษาศาสตร์และภาษาอังกฤษ Prof. Elena Semino อีเมล์ e.semino@lancaster.ac.uk หรือโทรศัพท์ +44(0)1524 594176

นางกมลชนก แสนเมือง (k.sanmuang@lancaster.ac.uk) นักศึกษาปริญญาเอก

ขอขอบคุณนิสิตที่เข้าร่วมโครงการในครั้งนี้

Appendix (6). Pre-test and Post-test

Instructions: There are three parts in this pre-test: choose the best groups of words (part 1), and in part 2 write up a paragraph by using the provided strings of words.

Part 1:

Instructions: From item 1 to 5, choose the appropriate strings of words.
(5 marks)

1. Alternating current (AC) is an electric current that repeatedly changes polarity from negative to positive and back again. The most commonly used form of alternating current does so in a sine wave pattern _____ Figure 1.

a. due to the b. because of the
c. as shown in d. figs show the

2. There is a more rapid rate of induction for 37 °C, _____high temperature, which causes a rapid build up and a lower plateau. Therefore, product inhibition occurs more quickly.

a. in order to b. due to the
c. in this study d. as well as

3. Therefore, _____ arrive at the point where a discussion of the strengths and weaknesses of the criticisms of critical realism is possible, one must first understand the argument that critical realism is positing, and so it is to chapter one that we presently turn.

a. in order to b. in terms of
c. in this study d. as well as

4. The graph showing _____ temperature on the product concentration shows a gradual increase getting sharper towards the peak at about 42 °C. After this point the concentration drops sharply as the temperature is further increased.

a. the use of b. the value of
c. the presence of d. the effect of

5. As _____large-scale water storage and distribution systems has increased, more buildings now contain reservoirs of L. pneumophila.

a. the number of b. the use of

c. the presence of d. the effect of

Part 2:

Instructions: From item 6 to 10, choose the correct groups of words in the box below to fill in the blank. (5 marks)

the number of	the effect of	it is based on the	the use of
due to the	the development of	the presence of	it can be

6. After this Nitrobenzene is reacted with concentrated sulphuric acid under reflux. Finally Thionyl chloride is reacted with the reaction mixture; SOCl2; then (NH4)2CO2, an Amine source. 10/12/05A2: Solubility product of Potassium Periodate Aim: To study the results of titration's with a sparingly soluble salt, and to investigate _____differing electrolytes. Theory: The equilibrium established between a saturated solution of a slightly soluble salt in contact with excess solid is a heterogeneous one.

7. Enzyme kinetics can also be described using the Michaelis-Menten equation. This involves_____ two parameters to describe the kinetic properties of enzymes - V max, the maximum velocity, and K M, the Michaelis-Menten constant. K M is related to the affinity of the enzyme for its substrate and is defined as being the concentration of substrate at which the velocity is at half its maximum value.

8. As the findings here are in line with the hypotheses made and in accordance with previous research and theory in the field, _____ concluded that results of this study are both reliable and valid, with generalisability of findings should further research be conducted along these lines.

9. This theory is predominately a psychodynamic one as _____ idea of 'reaction formation', as a forbidden impulse is re-channelled into a safer course. In effect these minority groups are used as scapegoats as they are considered to be weak and powerless.

10. Reliability factor would also increase if the precision of cutting is important. Developments could also be performed in order to improve the performance. Alternative materials could be used if the fatigue strength

was not sufficient enough. Increasing _____ shaft sample would also improve the reliability of the analysis, which may also increase the variety of effective design.

Part 2:

Instructions: Use the provided 'groups of words' below to produce a paragraph.

There are a few topics that you can write about such as, *the new technology used for saving power or environment, your engineering laboratory report, the weak points of having too advanced technology, future of technology adapted into smart homes etc.*

You should use at least 2 groups of words in the box to help write your paragraph, but no more than 5 groups of words. You can use your dictionary, if necessary. (10 marks)

the number of	the effect of	is based on the	the use of
due to the	the development of	the presence of	it can be
in order to	the value of	(as) shown in fig	as well as
(is) based on the	in this study	the presence of	in terms of
in this paper	as a result	the performance of	according to the
is used to	a function of	in this case	can be seen
one of the	the formation of	the stability of	compared to the

Appendix (7). Students' writing task 2

S1: K. U.

Pretest:

In technology design must have *the effect of* produce and protected system to best, future *the number of* many people everyone want have need in technology on intern swell up. I hope engineering everyone must design produce your best.

Posttest:

I have done about project changing temperature between heating and cooling. The effect of green house effect makes it earth warming. When is the temperature changes makes voltages Due to the I did a project temperature changes.

S2: J. C.

Pretest

My name is J. J. I was studying electrical engineering and computer at Kasetsart University. I'm doing a project about Solar cell. My project due to the environment of Thailand. I have done with my two friends. I decided to do this project.

Posttest: J. J.

My name is J. J. I study electrical and computer engineering. I have to do this one project. I have a computer to the use of the search for information. The number of friends that is in a group have skill on solar energy due to the sunlight is clean energy and in infinite energy. In order to analyse the effect of the sunlight, I used in the production of electricity in the home, it help save on energy costs.

S3: N. Y.

Pretest:

The development of the formation of in this paper a function of the performance of in this study

Posttest:

Hello, I'm N. Y. In order to provide convenience to a people, I am doing a project on the use of on-off smartphone. The smartphone is used to connect to a wifi.

S4: N. R.

Pretest:

The electrical engineering is difficult major. You'll be learn about electric current and It can not be seen. It can be learned by the imagination. On of the things that students need to meet the electricity is electrocuted.

Posttest:

My project is about robot. This robot is used to works with farmer, in order to help contain fruits. It can be automatic following. This robot can not contain over to 50 kilograms.

S5: P. S.

Pretest

In a study if wind energy, we can see that. At present, Thailand is not widespread Because of the high costs and in the area is limited is used to wind is less.

Posttest:

The power transformer have age *the use of* 20-30 years. The tank is Metal and Sealed,The brack down of transformer *due to the* PD in transformer .We can measure PD *In order to*protect the brack down. The effect of transformer is Discharge. The power transformer have age *the use of* 20-30 years. The tank is Metal and Sealed ,The brack down of transformer *due to the* PD in transformer .We can measure PD *In order to* protect the brack down .The effect of transformer is Discharge.

S6: P. P.

Pretest:

I want study of Kasetsart University, for the development if me. But the number of many students, makes me match with someone students. I want the formation of my family. So that, I study of Kasetsart University.

Posttest:

I do energy saving house model projects. I had to use plywood in the project, but if the number of plywood is very high. I was doing this project because we study the effect of global warming. I work on a project to conserve energy, due to the global warming everyday.

S7: V. K.

Pretest:

Solar cells for study the development of solar frame but can see how to keep underneath and the formation of solar cells that as a result how, Impressions from the number of surver results

Posttest:

I do learning potential electricity production from Solar cell project. In order to study the use of Solar cell, I analysis the performance of sunlight in Kasetsart University. I study the development of electricity from solar panels, for compared to the other renewable energy.

S8: W. S.

Pretest

The use of solar energy Instead of using the coal can be the amount of pollutants because many aspects such as the air landscapes

Posttest

Global warming is due to the ctivity of human, such as burning garbage and the smoke out of the car, etc. The use of plastic should be reduced. Therefore, we should plant more trees.

S9: S. K.

Pretest

The new of my produce is Energy saving lamp in this study can reduce Global warning as well as can reduce electricity charge.

Posttest

The number of Thai people use electricity is increasing due to the higher temperature. I have to save energy in order to reduce Global warming. The use of compact Fluorescent Bulbs can save electricity. The effect of Global warming can cause The end of the world.

S10: S.-U. K.

Pretest:

I project on electricity in building. The use of electricity in school building. There used to be many. I have installed solar cell the number of three kilowatt. To reduce the cost to the university

Posttest:

I am studying at Kasetsart University. My major is Electrical and Computer Engineering. The number of students is increasing every year. I have done a project about energy in order to reduce electricity costs in University. I did a project on energy due to the Energy used on Campus much higher. Currently, the use of solar array is very popular in Thailand.

S11: M. U.

Pretest:

Not written

Posttest:

Current, the number of car is increasing every year. cause global warming due to the carbon dioxide accumutate in the atmosphere.

S12: S.

Pretest:

I'm study electrical engineering due to the liked it and the effect of do it well but I'm not good English. I don't examination. Bye

Posttest:

I'm study electrical engineering, I study the use of electrical equipment in laboratory , the number of equipment so much and required equipment carefully. Due to the price of the expensive equipment and dangerous , may the effect of the body , and I study electrical engineering in order to keep the family as well when I have work to do. Thank you.

S13: A. H.

Pretest:

Solar energy is heat energy from sunshine manufacture electricity one way to reduce the effect of not Earth and manufacture electricity the performance of is one option.

Posttest:

OHMASA_GAS a fluid stirrer that does not is used to mechanical vibrations of low frequency instead. The use of water circulation has good quality. If you pour detergent into this water, bubbles son't occur. Ordinary electrolysis makes O_2 and H2 gas. Tis low frequency vibration breaks water surface tension, so bubbles can't occur is electrolysis.

Appendix (8). Materials piloting results

Overall, I have distributed 11 copies of the materials for the trial (9 teachers and 2 students in the field of science and engineering). There are some good points and issues that need to be reconsidered carefully as follows:

Good points:

Activities:

9 out of 10 participants thought that the ways the activities have been organized seem to be okay. For example, one of them agreed that the logic of the activities can promote 'analysing skills'. The self-study section is quite helpful for them to remind them of the language points (subjects and objects).

In my opinion, I think this can help them to increase the awareness of the use of subjects and verbs in the texts. However, there are still some issues that need to be considered in order to produce the 'ideal materials' for this specific context.

Issues need to be considered

1. Theory of teaching: One of them has raised the point 'should teaching theory be involved in the materials'.

2. Objectives: 2 of the pilotors agreed that I should provide the *main objective* in the very beginning of the materials, the *focus of the teaching,* and *the overview of the teaching* should be informed explicitly.

3. Language use in the materials:

Some participants felt that the materials were quite challenging for engineering students due to the use of linguistics terminology, such as the phrase lexical bundle. Additionally, certain vocabulary, like dependent complement clause and prepositional in Activity B, was considered too technical for this audience. To address this, I should simplify the language and make technical terms more accessible and easier to understand for engineering students. One of them thought that the students might get difficulty when studying this in class, but if I can facilitate them, they can be able to do the exercises.

- Language use in the instructions: The word 'notice' in activity C should be replaced by the words 'pay attention to/study'. BUT I think the word 'notice' lead the students to explore the lexical bundles themselves.
- The students might get confused with some questions and words. For example, I should make sure what I mean by the 'target word'.
- The word '*potential meaning'* in Activity 2 can make the student confused. I should better say *'is important commonly used in positive or negative context?*
- The questions, *What is the verb that is always used before 'important'? and What are the other types of*

words used before 'important'? can hinder students' understanding, specifically the words '*types of words"*? The engineering students seem to fail to understand what *'the types of word'* are.

4. Activities:

One of the pilotors is a PhD student in engineering who found that the materials is indeed helpful for him in improving his research articles writing. However, he suggested that I should revise the activities in part 1 (how to use AntConc) in a very detailed because the students might not have confidence when they cannot use the programme. He suggested some useful improvement on his own view as a teacher and as a student as following suggestions:

- He suggested that the 1L (one-word-to-the-left),1R (one-word-to-the-right) should be explained clearly. Or can they only search only 1L, 2L, or they can search 1L and 1R together.
- KWIC sort section: level 1, level 2, level 3 should be introduced to the students.
- I should clarify what does it mean by *'it looks okay'* on page 6 (*as you can see that it looks okay and you are now ready to do a search within the concordance tool*).
- The instruction in Section 1 should be in affirmative form, and provide step-by-step explanation because it might cause techno-phobic for the students who cannot follow the instructions, and cannot use the programme skillfully.
- I should divide activities into specific engineering discipline; for example, the lexical bundles that occur frequently in one discipline should be used to create activities in that discipline. I should provide choices of activities in the 4 disciplines and allow the students to work in their own disciplines (write their own paragraph). The main reason is that the students know best in their own field and can help them to write better in their specific field.
- Alternatively, I can choose the basic foundation subjects that all students have been already studied such as, basic engineering, mechanics, physics and then later separate the students to work on the activities according to their disciplines.
- The concordances I selected to teach should not be longer than 10 lines because the students' level of proficiency.

Sequences of the activities:

- Some activities should be rearranged for example; I should introduce the concordances and node words before introducing the AntConc. **BUT** I think the present sequence is okay because the students need to gain confident in using the technology before they play with the concordances and node words.
- In the section '**Searching for a word and seeing its concordances'**, I should not start with the word **'**the number of**'** because it might be too difficult for students to pick it up**.** One of the pilotors suggested that it is better to start with something easy like asking the students to examine the word 'house'. So, I can point

put to the students *what the node is* and *what can be explored* from the concordance lines, or *what techniques* they can use when playing with concordance. Then, I can give them a few minutes to play with the tool. Allow the students to search any word they want.

5. Clarity of the instructions: the instructions should be clear and more specific

- For example, the instructions in activity B, *The following patterns tell us about how the lexical bundles (the effect of, the number of) are used in the research articles. Can you group the right concordances with the appropriate pattern? How?* I should explain more or give some examples.
- The instructions in activity C, *Identify what types of word usually follow them such as, nouns, verbs*, collocations, adverbs. The technical words, *collocation, adverbs, nouns, verbs,* should be told explicitly.
- Also, the instructions in activity C, *Do you think the words that co-occur with the lexical bundles mentioned play an important role in the sentence, and indicate specific actions in engineering?* I should make these questions more specific and be able to point out what kinds of words I want them to pay attention to, and how? Those questions are quite hard for engineering students to be able to answer.

6. Others points:

- some grammatical mistakes
- formatting: the concordances of the important in Activity 2 should be centered.

Appendix (9). Judgements of writing activities from the three raters: Pre-test & Post-test part 2

Pre-Test

Student No.	Pre2 scores	Correct	Incorrect	Rater 1 correct	Rater 2 correct	Rater 3 correct	Coding	Rater 1 incorrect	Rater 2 incorrect	Rater 3 incorrect	Coding
S1	0	0	2	0	0	0	1	2	2	2	1
S2	0	0	1	0	0	0	1	1	1	1	1
S3	0	0	0	0	0	0	1	0	0	0	1
S4	0	0	1	0	0	0	1	1	1	1	1
S5	0	0	1	0	0	0	1	1	1	1	1
S6	0	0	2	0	0	1	0	2	2	1	0
S7	4	2	1	2	2	2	1	1	1	1	1
S8	2	1	0	1	1	1	1	0	0	0	1
S9	0	0	2	0	0	0	1	2	2	2	1
S10	2	1	0	1	1	1	1	0	0	0	1
S11	0	0	0	0	0	0	1	0	0	0	1
S12	0	0	2	0	0	0	1	2	2	2	1
S13	0	0	0	0	0	0	1	0	0	0	1
	Percentage of agreement						92.31%				92.31%

Post-Test

No.	Post2 scores	Correct	Incorrect	Rater 1 correct	Rater 2 correct	Rater 3 correct	Coding	Rater 1 incorrect	Rater 2 incorrect	Rater 3 incorrect	Coding
S1	2	1	0	1	1	1	1	0	0	0	1
S2	6	3	2	3	3	3	1	2	2	2	1
S3	4	2	0	2	2	2	1	0	0	0	1
S4	4	2	0	2	2	2	1	0	0	0	1
S5	6	3	1	3	3	3	1	1	1	1	1
S6	4	2	1	2	1	1	0	1	1	1	1
S7	6	3	1	3	3	3	1	1	1	1	1
S8	4	2	0	2	2	2	1	0	0	0	1
S9	10	5	0	5	5	5	1	0	0	0	1
S10	6	3	0	3	3	3	1	0	0	0	1
S11	0	0	2	0	0	0	1	2	2	2	1
S12	8	4	0	4	4	4	1	0	0	0	1
S13	2	1	1	1	1	1	1	1	1	1	1
	Percentage of agreement						92.31%				100%

Appendix (10). Judgements of grammatical errors that are related and unrelated to the use of the lexical bundles in students' writing: Pre-test & Post-Test Writing activity 2

Students' writing in the pre-test

Students' writing in the pre-test

No. of sentences	St. No.	Grammatical errors: Sentences that are related to lexical bundles	Grammatical errors categories: Errors within the bundles	Grammatical errors categories: other errors	Meaning
1	S1	In technology design must have **the effect of** produce and protected system to best, future **the number of** many people everyone want have need in technology on intern swell up.	verb form/noun phrase form (many people)		yes/no
2	S2	My project **due to th**e environment of Thailand.		verb missing/wrong word choice	no
3	S3	Not written anything			
4	S6	I want study of Kasetsart University, for **the development of** me.	the bundle itself is a wrong word choice		yes
5	S6	But **the number of** many students, makes me match with someone students.	noun phrase form		no
6	S6	I want **the formation of** my family.	the bundle itself is a wrong word choice		no
7	S7	Solar cells for study **the development of** solar frame but can see how to keep underneath and **the formation of** solar cells that as a result how, Impressions from **the number of** surver results	doesn't make any sense at all		no/no/no
8	S8	**The use of** solar energy Instead of using the coal can be **the amount of** pollutants because many aspects such as the air landscapes	the grammar is fine before **the amount of**, but after that the thing doesn't make any sense		yes/no
9	S9	The new of my produce is Energy saving lamp **in this study** can reduce Global warning **as well as** can reduce electricity charge.	subject missing		yes/yes
10	S10	**The use of** electricity in school building.	no errors related to **the use of** but the whole thing is a fragment		yes
11	S10	I have installed solar cell **the number of** three kilowatt.	wrong word choice (the number of)		no
12	S11	Not written anything			
13	S12	I'm study electrical engineering **due to the** liked it and **the effect of** do it well but I'm not good English.	noun phrase form/after "the effect of" doesn't make any sense		yes/no
14	S13	Solar energy is heat energy from sunshine manufacture electricity one way to reduce **the effect of** not Earth and manufacture electricity t**he performance of** is one option.		the whole thing doesn't make any sense	no/no

No. of sentences	St. No.	Grammatical errors Sentences that are NOT related to lexical bundles	Errors	Grammatical errors categories other errors	Meaning
1	S1	I hope engineering everyone must design produce your best.			No
2	S2	My name is J. J.			
3	S2	I was studying electrical engineering and computer at Kasetsart University.			
4	S2	I'm doing a project about Solar cell.			
5	S2	I have done with my two friends.			
6[23]	S2	I decided to do this project.			
7	S4	The electrical engineering is difficult major.	article		Yes
8	S4	You'll be learn about electric current and It can not be seen.	verb form	spelling	Yes
9	S4	It can be learned by the imagination.			
10	S4	On of the things that students need to meet the electricity is electrocuted.		the whole thing doesn't make any sense	No
11	S5	In a study if wind energy, we can see that.	fragment		No
12	S5	At present, Thailand is not widespread Because of the high costs and in the area is limited is used to wind is less.		the whole thing doesn't make any sense	No
13	S6	So that, I study of Kasetsart University.	connector/preposition		Yes

[23] The highlighted sentences refer to correct sentences.

	S1		word choice		
14	0	I project on electricity in building.	(verb)		Yes
	S1		fragment/preposition		
15	0	There used to be many.			Yes
	S1		fragment		
16	0	To reduce the cost to the university			Yes
	S1				
17	2	I don’t examination. bye		verb missing	No

Students' writing in the post-test

No. of sentences	St. No.	Grammatical errors Sentences that are related to lexical bundles	Errors within the bundles	Grammatical errors categories other errors	Meaning
1	S1	**The effect of** green house effect makes it earth warming.	I think the use of LB is correct		yes
2	S1	When is the temperature changes makes voltages **Due to the** I did a project temperature changes.	wrong choice of bundle		no
3	S2	The **number of** friends that is in a group have skill on solar energy **due to the** sunlight is clean energy and in infinite energy.	subject and verb agreement/wrong choice of LB (doesn't make sense)		yes/no
4	S2	**In order to** analyse **the effect of** the sunlight, I used in the production of electricity in the home, it help save on energy costs.	the first chunk that consists of the bundled are correct		the second sentence doesn't make sense
5	S2	I have a computer to **the use of** the search for information.	wrong choice of bundle		yes
6	S2	**The number of** friends that is in a group have skill on solar energy **due to the** sunlight is clean energy and in infinite energy.	same as number 6		
7	S3	**In order to** provide convenience to a people, I am doing a project on **the use of** on-off smartphone.	the use of the two LBs are correct		yes
8	S3	The smartphone **is used to** connect to a wifi.	the use of the bundle is correct		yes
9	S4	This robot **is used to** works with farmer, **in order to** help contain fruits.	verb phrase form/correct		yes
10	S5	The power transformer have age **the use of** 20-30 years.	wrong choice of bundle		no
11	S5	The tank is Metal and Sealed, The brack down of transformer **due to the** PD in transformer .	main verb missing		yes
12	S5	We can measure PD **In order to** protect the brack down.	the use of LB is correct		yes
13	S6	I had to use plywood in the project, but if **the number of** plywood is very high.	not sure what she means by "the number of"		
14	S6	I was doing this project because we study **the effect of** global warming.	correct use of LB		yes
15	S6	I work on a project to conserve energy, **due to the** global warming everyday.	correct use of LB		yes
16	S7	In order to study **the use of** Solar cell, I analysis **the performance of** sunlight in Kasetsart University.	correct use of LB		yes
17	S7	I study **the development of** electricity from solar panels, for compared to the other renewable energy.	not sure what she really wants to convey		
18	S8	Global warming is **due to the** activity of human, such as burning garbage and the smoke out of the car, etc.	wrong choice of bundle		yes
19	S8	**The use of** plastic should be reduced.	correct use of LB		yes
20	S9	**The number of** Thai people use electricity is increasing **due to the** higher temperature.	correct use of LB		yes
21	S9	I have to save energy **in order to** reduce Global warming.	correct use of LB		yes
22	S9	**The use of** compact Fluorescent Bulbs can save electricity.	correct use of LB		yes
23	S9	**The effect of** Global warming can cause The end of the world.	correct use of LB		yes
24	S10	**The number of** students is increasing every year.	correct use of LB		yes
25	S10	I have done a project about energy **in order to** reduce electricity costs in University.	correct use of LB		yes
26	S10	I did a project on energy **due to the** Energy used on Campus much higher.	due to the + noun phrase		yes
27	S10	Currently, **the use of** solar array is very popular in Thailand.	correct use of LB		yes
28	S11	Current, **the number of** car is increasing every year. cause global warming **due to the** carbon dioxide accumutate in the atmosphere.	form of noun after "the number of"/due to the + noun phrase		yes/yes
29	S12	**Due to the** price of the expensive equipment and dangerous , may **the effect of** the body , and I study electrical engineering **in order to** keep the family as well when I have work to do.	correct use of LB/doesn't make sense/correct use of bundle but wrong choice of verb after "in oder to"		yes/no/no
30	S12	I'm study electrical engineering, I study **the use of** electrical equipment in laboratory , **the number of** equipment so much and required equipment carefully.	correct use of LB/noun after "the number of"		yes/no
31	S13	OHMASA_GAS a fluid stirrer that does not **is used to** mechanical vibrations of low frequency instead.	doesn't make any sense		no
32	S13	**The use of** water circulation has good quality.	correct use of LB		yes

	1. Pakpoom	explain briefly what happen?How does the ss response?	verbs used	Code	Main themes
1	M: What do you think about the training sessions? (Do you like it? How useful	I have asked him about the sessions. Whether			my thinking
2	is it to you?)	he likes it or not?			
3	P: I think I can apply the use of lexical bundles in writing documents in English	He replied that the LBs help him to write	think	Yes, helpful	
4	for example, if I would have asked to write English document in English. This is	English documents. He responded that the		consult with examples of the language	
5	the first time I have known that there is a programme like this, I can see many	examples showned in AntConc help him to know	have known	used in the programme	
6	examples from the programme especially, the specific words that are used in	how the specific words are used in the Ras.	first time	new knowledge	
7	my engineering field (electrical engineering).	He will use LBs in his writing.		for example, the students did not	
8	M: The lexical bundles (LBs) that you have been learnt in the programme i.e.			know the meaning of house in the	
9	using concordance lines, do you think they are useful?	He responded that the LBs are useful		concordance lines in the classroom	
10	M: I think it is useful. I think I will have to use a lot of them.		think, think		
11	M: What aspects of the training sessions do you like most (e.g. learning	I asked him what does he like most in the sessions.	will have to		
12	materials, content, etc.)? Why?				
13	P: I like the ways you taught us and the programme. As I have said previously	He likes the teaching methods, using	like	new knowledge	
14	that I have never seen this programme before.	concordance lines, and the programme	have never seen	technology	
15	M: How about the materials?				
16	P: I think it is good, I had to think a lot while I was doing the activities. I have	his attitudes toward the materials is good	think		
17	to think a lot and I can revise them later. I normally did not look at it.	he agreed that the materials encourage thinking skills	have to think a lot	thinking skills	
18	M: Do you mean you do not look at the materials? Do you think they are	Even though he said the materials are good but	revise	used to revise	
19	helpful? Any points that you think we could make the materials better?	he rarely reviews the materials. He thinks the	did not review	materials should be something	
20	P: I did not review them much, but I do not think you should revise them.	materials should not be revised.	did not think	that students can use only in the	short but effective
21	They are already good.		good	classroom	
22	M: How about the contents in the materials?	in terms of contents			
23	P: I think they have covered everything. They have a lot of vocabulary.	on his own opinion, the materials have covered	think	vocabulary	
24	M: How useful are the materials?	everything, especially the vocabulary			
25	P: I think they are useful, I have seen new words. If any words that I am	He thinks the vocabulary is useful for him	think, have seen	interesting words used in the engineering field	
26	interested in, I will translate them and I know that those words are words	he will only translate the words that have caught	will translate	translation' strategies when the words are of	
27	that are used in my engineering field.	his attention. But only the words are used in the	know	interest	
28	M: How useful is the teaching of lexical bundles (fixed sequences of words)?	electrical field.		the words are interesting and meaningful because	acquisition can happen if he has existing
29	P: The teaching of lexical bundles is absolutely useful for me (with very high	He responded to the usefulness of the LBs in	absolutely	they are in the engineering field and he uses the existing	knowledge
30	degree of confidence)	the absolute sense	useful	knowledge or the contents to get a new knowledge	LBs have links to the specialised field on ways or another

Appendix (K). Samples of Students' interview and teacher's note

	1. Pakpoom	explain briefly what happen?How does the ss response?	verbs used	evaluative verbs	Code	Main themes	Sub-themes
31	M: In what ways, the teaching of LBs is useful?	I have asked him about the ways of teaching LBs				It has been confirmed by other students	specialised word to LBs
32	P: I think it is useful when I have to work on the project, help in learning, and	in the classroom, giving lots of example, searching	think, useful			when they search the words they are	
33	know the vocabulary.	for the patterns	have to work			interested in and it linked to the taught LBs	
34	M: Only know the vocabulary? Do you usually learn only an **individual word**,		on the project				
35	but I have taught you group of words.		help in learning				
36	M: To what extent do you think the ERAs sessions help improve your English skills?	He responded that his grammar skills have been			lexico-grammatical		
37	P: I think now I know what is an adjective.	improved. He give an examples 'the adjective'	think, know				
38	M: How?						
39	P: I try to **notice t**he words that occur after or before that words.	He responded that to be able to know the adjectives	try to notice		first attempt to notice the adjective used in the concordance lines. The noticing suits his learning styles	encourage noticing skills	DDL
40	M: And in what aspects (e.g. words chunk, grammar, syntax, semantic)?	he attempted to notice the words that occur before					
41	P: I think I feel more confidence with my English now. To be honest, before	or after the the target LBs. Then, he knows	feel confidence		a sense of accomplishment raises his confidence	corpus plays an important role in students	noticing and acquisition
42	coming to the session, I think my English is very poor and it is not going to be	what are the adjectives. He has taught that his English	think		evaluate his English ability	learning strategies, noticing	motivation can make him acquire
43	improved.	skills are poor, but he feels confident after attending					the language
44	M: Am I right to say that 'the sessions have changed a lot of your thinking about English.	the session.					
45	It makes you think that you are able to learn English. You get more	I have emphasized that the sessions have encouraged him	are ble to learn more confidence				
46	confidence.	to think a lot and increased his confidence	English				
47	P: Yes, you are right.						
48	M: What else?	I have continued to ask him on which aspects from					
49	P: structure, nouns, adverbs, adjectives	the sessions that can help improve his English.	grammatical points				
50	M: You see the nouns, adverbs, and adjectives from where, from loads of	He responsed that he can notice the mouns, verbs, adjectives	seen		visual sense		
51	examples we were learning together?	from examples provided from concordance lines.					
52	P:I have seen them from loads of examples in the concordance lines and from	He emphasied on the loads of example that I have taught	have seen				
53	what you have been teaching us	in the programme.	loads of examples				
54	M: How about the concordance line? Do they provoke your thought? meaning to	I have asked whther the concordance lines have effects					
55	say that when you see the concordance line, you can notice which words	the noticing skills. Whether they have related to each other					
56	should be followed by which words	in his own view.					
57	P: they are very useful and when I see them I know that those lexical bundles	He responded that they are related and it enables him to	see/know				
58	words should mainly follow nouns. It they are in the very beginning of the sentences,	identify the noun and so on.					
59	they should be followed by verbs and nouns.	He explained his ways of noticinng					
60	M: so in what aspects, do you think the training sessions help improve your						
61	English skills ?						
62	P: I think vocabulary (1), and my attitudes towards English as I previously told	He said that the training sessions help improve the lexical	think		needs and lacks (Jordan, 1997)	strategy analysisi	
63	you that I am considered as 'very poor English user', BUT I think I have been	skills, by refering to his attitudes towards his English ability	consider		students'utilise learning strategies	initiated the fluency at early stage of acquisition	
64	improved my English such as nouns (1), specific words (2) used in my	previously. He considered his ability as 'poor'. The attitudes has changed	think		promote a sense of achivements		
65	engineering field. and (3) structures	after the session, he has gaine more confidence in					
66	Do you think you can extend the knowledge of using LBs in your future, based	his English ability because he has achieved the activities.					
67	on what you have been trained?						
68	P: Yes, I think I can apply the knowledge in the writing of my project and		think/can apply				
69	by using the AntCnc programe.						
70	M: How about using in writing the engineering research articles?						
71	P: Yes, I think so. There are examples from the concordance lines and I can look		think so				
72	and analyse the patterns of those lines						
73	M: How about the contexts? the words that co-occur with the target word, have						
74	you realized this issue?						
75	P: Yes, I have, but not all the times						
76	M: Do you have any comments on the training sessions?						
77	P: No, I think the session is good.		think				

	2. Sakowduen	explain briefly what happen?	verbs used	Code	Main themes	Sub-themes
78	M: What do you think about the training sessions? (Do you like it? How useful	I have asked her about the sessions. Whether				
79	is it to you?)	she likes it or not?				
80	S. It was fun but I think it all depends on my **English background**	She enjoyed the session	was fun	refer to English proficiency: low	students'identify	
81	and some words I still did not know the meaning of them.	She responded that she did not know the meanings	did not know	lack of strategies		
82	M: What English classes you have studied ?	of some words. So, I think I should ask her background		new knowledge		
83	S: Umm **English conversation, English structure, but I forget all of them**.	about her previous English courses she has studied.		previous course, only grammar		
84	M: **Have you reviewed** all of them, right?	I have asked what English courses she has studied		but no strategies in writing ERAs	strategies	writing strategies for
85	S. **No**, I have not reviewed all of them at all.	because she has passed all English foundation, and		have not reviewed the previous knwledge		LBs
86	M: How useful is the teaching sessions?	then she should have skills to be able to find the meanings in the context	have been learning	learnt the engineering words in Thai	memory strategies in LB	
87	S. I have been learning **engineering words in Thai**. I have not never studied the	of the words in the ERAs.	never studied	have no background in what she is studying		
88	articles in English. First, I felt that those words are **new** to me, but with **your help**	I asked about the usefulness of the sessions.	Felt	words are new	teachers role	faciliate
89	trying to **explain** to me.	She accepted that this has been the first time that she learnt the words	with your help/explain	explaination from the teacher		
90	M: Do you think you have exposed more English words in your field in the sessions?	in English and she found it difficult. However, the support from the teacher	Like			
91	S. Yes.	is important for her to understand what she is learning.	have never seen			
92	M: What aspects of the training sessions do you like most (e.g. learning	I asked her about what aspects of the training sessions she likes most?				
93	materials, content, etc.)? Why?	She agreed that she liked the use of LBs and with the concordance lines				
94	S. I like the ways you taught us. You have not only **explained on the points that I have**	in helping her to be able to get the meanings from what are around the context.	like			
95	**confused or could not follow,** but you have also encouraged us to think first. In another	The support from teacher, explaination is very important. The support here means	have confused/could not follow	support from teacher		
96	words, **you did not explain us straight away, but I have to think it through first** before	the teacher did not fully help her to get the answer. She had to try thinking on	have encouraged to think			
97	I have been given the answer. In the classroom, I get used to **the spoonfeeding,** the	her own first. She used to be given the answer quickly after the questions asked	did not explain the answer			
98	teachers just **gave me the answer right away**. I did not have a chance to think.	by the teachers.	think it first. Have no chance to think			
99	In the sessions, I had to **think a lot**, and **do the tasks myself**. I am able to **remember and**	She agreed that thinking and practice on the tasks have helped her understand	think a lot			
100	**understand** the contents.	the LBs.	do the tasks myself	promote independent		
101	M: How about the **examples of the use of LBs** in the ERAs you have seen the concordance lines					
102	To what extent do you think these help?					
103	S. the examples right. I think they are **very helpful**. For example, I have to analyse line by line	She has valued the concordance lines, she can analyse the concordance lines in	think	concordance line		
104	and can look at the original files (the ERAs). I have looked at the subjects or objects and now	small amounts, by looking at the functions of the words, and whether they are	helpful			
105	I think I am able to distinguish which is subject and which is object, and with the help of the	subject or objects, including the help of AntConc.	analyse, look	increase ability in identify subject and		
106	AntConc.		think I am able	object		
107	M: How useful are the materials?					
108	S. It is indeed useful. I can use this materials to review when I have to do the writing tasks.	She said that she can use the materials to review the contents, and when	indeed	emphasize the usefulness of the materials		
109	For example, when I was assigned to do homework, I can go back and revisit what I have been	writing, revisit				
110	learnt.					

111	M: Have you ever experienced this kind of learning?	I have asked her whether she has familiar with this learning.				
112	S: No, this has been the first time I have experienced this. I have studied basic English.	She said that it has been her first time.		first time		
113	M: Do you think the sessions should be taught to your junior?					
114	S: Yes, I think it is a good course because we all have been studied only general English.	She responded that the course is good and provide more than general English		think		
115	I think this course seems to be **another step in learning English in our engineering field**	and she said that the course can help with English in her engineering field.		think see know		
116	that could equip with the writing			equip with writing		
117	M: Do you think the teaching fits well with your field (electrical engineering)?					
118	S: I think it is very helpful and useful.	She responded that the teaching is very helpful and useful.		helpful and useful		
119	M: How useful is the teaching of lexical bundles (fixed sequences of words)?		I asked her about the LBs teaching.			
120	S: I think it is totally helpful because normally I **pay attention to only an individual word**		She agreed that the teaching of LBs is useful for her because she had a chance to	think		single word and bundles
121	and only some words and all are in a single word. I have never seen the words that come		compare the single word and the strings of words.	pay attention to a single word		existing learning strategies
122	together in groups like the 3-word bundles that you have introduced. I think they are useful		She expected to see the 3-word bundle in her field	think		
123	and I think I expect to see them again in my study in engineering.			expect to see LBs		students' attitudes towards LBs
124	M: To what extent do you think the ERAs sessions help improve your English skills?	I asked her the English skills she thinks she has improved.				
125	And in what aspects (e.g. words chunk, grammar, syntax, semantic)?					
126	S: my vocabulary, loads of them in the sessions, structures, patterns of those words	vocabulary		think		
127	M: can you give me examples of the words you learnt and to what extent the LBs help improve					
128	your vocabulary?					
129	S: Not only LBs that I have learnt but I did have t**o notice** the words that come before or after			think so		
130	the LBs. I do not normally pay attention to the words that come before or after the LBs.					
131	However, I did pay attention to their meanings, types of words.					
132	M: How about the structure?					
133	S: Noun Phrase (NP), I never know what the NP is. This has been the first time.					
134	M: Anything else?					
135	S: The subjects and objects that are used with the LBs.			think		
136	M: Overall, do you think the training sessions help improve your ability to write					
137	articles in engineering?					
138	S: It helps me a lot. When I write in English, I am aware of the meanings of the words, and					
139	which words should be used with which words, such as nouns, adjectives, verbs. Therefore,					
140	I can use the strategies in my day to day writing.					
141	M: Are you going to use the LBs in writing some sections in the Ras?					
142	S: I think I will use the LBs in the other courses that use English as a medium of instruction.					
143	I will us the LBs in writing job application letter, sentences construction and write my CV					
144	M: Do you have any comments on the training sessions?					
145	S: I think your session is good already, only concerning the time allocation.					
146	M: Do you think the time allocation for the session should be longer?					
147	S: No, I think the time allocation is okay.					
148	M: You know 'how to', then you can apply in your learning.					
149	S: Yes, your teaching is very good, you are very easy-going You can make us all understand					
150	you know not everyone who attended the sessions are good at english, but they seem to					
151	understand the contents.					

	3. Kittichai (attended some sessions)	explain briefly what happen?	Code	Main themes	Sub-themes
152	M: What do you think about the training sessions? (Do you like it? How useful				
153	is it to you?)				
154	K: I like it, it is useful, **in particular the programme** I know how to use it.		refer to English proficiency: low		
155	M: Only the programme?		not confidence		
156	K: There is also LBs, I am NOT good at **vocabulary**. I **do not like English** because I have no		new knowledge		
157	ideas of what are the **subject, verb, or object**.				
158	M: You have **already passed English** foundation I - III. Why you said you **do not know**		programme		
159	subject, verb, or object?				
160	K: I **tried to make sense** of what the teachers taught in classroom, and **utilised the knowledge**				
161	**in the exam. I always memorise the words and the meaning**				
162	M: What do you think about the sessions?				
163	K: I think I did not understand. I have no ideas of for example what is verb, adverb or noun ?				
164	M: You only attended some sessions!				
165	K: I think my **English background is not good**. I do not even know adjective, verb for example.				
166	M: What aspects of the training sessions do you like most (e.g. learning materials, content, etc.)?				
167	Why?				
168	K: The ways how to **NOTICE** the words in the programme. I have been taught 'the sentences				
169	that come after'. I do not like English especially in the texts. I used to do the 'error' and 'essay'				
170	exams, but I did not understand. After studying in this sessions, I think I can understand 30%, in				
171	the ways that I should notice parts in the research.				
172	In the sessions, I had to **think a lot**, and **do the tasks myself**. I am able to **remember and**				
173	M: How useful are the materials?				
174	K: Yes, I think I can use it to study after the sessions.				
175	M: How useful is the teaching of lexical bundles (fixed sequences of words)?				
176	K: Yes, it is. If we notice each part of the RAs, each part can have LBs and the words that follow				
177	whether what are the words and the meaning				
178	M: To what extent do you think the ERAs sessions help improve your English skills?				
179	K: I think structure, noticing for example if I find the LB, the words that come before or after				
180	should be like this.				
181	M: The words that I have taught you that come before or after or co-occurr with LB.				
182	Does it help you to improve your English?				
183	K: yes,I can know the words that can come after.				
184	M: How about the concordance lines?				
185	K: I can study on my own and from the use of the programme				
186	M: So, when you went back to the orginal file, you will see the Ras such as the words that are used				
187	in your engineering field? Overall, do you think the training sessions help improve your				
188	your ability to write articles in engineering?				
189	K: waiting time... Nouns I think I should know nouns and grammar				
190	Noun Phrase and I can utilise them, NOTICING I can remember it for long time.				
191	M: Actually, in the sessions I did not intend to teach you the memorising, instead encouraging				
192	awareness of the LBs used in the RAs. Am I right to say that you remember which LBs should use				
193	with which structures?				
194	K: Do you have any comments on the training sessions?				
195	K: I think it depends on the students really. The students are required to attend all the sessions				
196	The students should be knowledgeable in the basic English , and can study English on their own.				
197	You (me) are very nice and I can get new knowledge.				

	4. Maruet	explain briefly what happen?	Code	Main themes	Sub-themes
199	M: What do you think about the training sessions? (Do you like it? How useful is it)				
200	M: I **like it but** my English is weak. I want to be knowledgeable in English, and want to extend		refer to English proficiency: low		
201	it. I think I should use it often. I was confused with the sessions in the very first time.		not confidence		
202	And then I know it is **one of the ways** of learning English that I never learnt before.		new knowledge		
203	Me: Can you tell me which ways that you refer to?				
204	M: I know the use of t**he words, and LBs** should be used with what. And the programme itself.		words/LB/sprogramme		
205	Me: What do you think about the **words and LBs**?				
206	M: I think the LBs should be used in my electrical engineering field, I think they are		words reflects to his future work		
207	specialised words.				
208	Me: What aspects of the training sessions do you like most (e.g. learning materials, content, etc.)?				
209	Why?				
210	M: I think the ways you have taught us, first you are more friendly, the materials is also okay				
211	there is a struction to follow of how to do, and the explaination. But it all depends on me,				
212	I have less experience in LBs.				
213	Me: How useful are the materials?				
214	M: The materials is very useful, I can follow the materials				
215	Me: You just looked at the documents and can do the activities?				
216	M: Yes to some extents				
217	Me: To what extents, the materials should be improved?				
218	M: You should separate the sections such as the use of the programme and the teaching.				
219	Me: How useful is the teaching of lexical bundles (fixed sequences of words)?		previous knowledge of LBs		
220	M: Yes, it is. I do not normally have knowledge in LBs. I am not even aware that what is LBs.		awareness of LBs		
221	I always use only individual words. You have taught me the LBs and the meanings.		previous knowledge of individual words		
222	Me: To what extent do you think the ERAs sessions help improve your English skills?				
223	M: It opens the door to me. English has more to larn than we thought.				
224	Me: How?				
225	M: Normally, I study grammar, learn only individual words. Studying LBs, I know that the LBs		LBs and context		
226	change through context. One word can combine with another word and we can get the LBs.				
227	Me: How?				
228	M: For example, the LB 'the effect of' is from the combination of three words and 'due to the'		relaisation the usefulness of the LBs		
229	as well. The individual word is always on its own. You have taught us 3-word bundle.LB				
230	learnt.				
231	Me: Do you think studying words in bundles or an individula help you learn English?		context and LBs		
232	M: I feel a bit confused of the LBs than the single words. Then, after I have learnt the LBs, I know		individual words VS LBs		
233	that it all depends on the **context.** The single words you can look at the dictionary.				
234	Me:do you think the training sessions help improve your ability to write articles in engineering				
235	M: I think I can have less chance to use LBs				

236	Me: Do the LBs help in the RAs writing?	
237	M: I have not found the chance to use them yet. But, if I know when to use them. I can use	
238	them of course.	
239	Me: How about using LBs in your abstract writing in your project?	
240	M: Normally, I have someone to do the English version for me such as friends, Google	use another shortcut ways to do the project
241	I mean friend who are good at English.	support for them
242	Me: But supposed you will not use the shortcut ways like that	
243	M: I think I will not use the Google translation of friend service, if the words are not the LBs	consider LBs as easy
244	I did not fully make sense of the LBs.	other words as easy
245	Me:Do you have any comments on the training sessions	translation into Google/friends
246	M: I think you should translate for me because sometimes I don't understand.	get used to the translation-learning
247	But it all depends on my **background,** I have a weak background in English, such as vocabulary	refer to English background
248	I want to use it in my daily life.	
249	Me: Do you think you will use the LBs in your daily life?	
250	M: Yes, I think it can help. But it depends more on the students'experience in LBs.	

	5. Waravut & 6. Nattakarn	explain briefly what happen?	Code	Main themes	Sub-themes
251	M: What do you think about the training sessions? (Do you like it? How useful				
252	is it to you?)				
253	N: It is okay. I think I know a lot. I have known what I never known before.		previous study		
254	Everything for me is totally new!				
255	W: I consider the LBs as **advanced skills,** such as **nouns+verbs+adverbs**		what students perceived LBs		
256	I have never learnt it before. I have less knowledge in LBs and my background		refer to English proficiency: low		
257	is not very good. **It is more difficult** when I have to study the LBs, because they are		what students perceived LBs		
258	advanced and I am not able to do it.				
259	M: Do you like it? What do you think? Does it helpful?		sessions attitudes		
260	N: It is indeed useful.		session is useful		
261	W: I think it is very useful because you taught me **something in the engineering field**		related to engineering field		
262	and we can use in **our expertise in which English is used as a medium**		related to their expertise		
263	M: Do you like it?				
264	N+M: Yes, we did like it, especially you are very friendly, we feel comfortable.		classroom atmosphere		
265	You know the contents is hard enough and if we are under pressure, it makes the lesson		students attitude towards English		
266	boring. We do not normally like English and in the sessions we did feel comfortable.		motivation before learning		
267	M: What aspects of the training sessions do you like most (e.g. learning materials, content, etc.)?		filter if ss don't feel comfortable		
268	Why?				
269	W: I like the materials, techniques in the programme, just types the target words. And then				
270	I can see the sentences, easy. We can see the patterns such as the use of + NP. Also we know				
271	theat the sentences are in the beginning part or middle part of the RA				
272	we cannnot see these in the dictionary				
273	M: What else?				
274	W: I think the materials is not as interesting as the programme				
275	M: To what extent do you think the programme help you to learn?				
276	W: I think your teaching is also important. Normally, we do not review what we have been				
277	learnt. You have explained us and friend too. I am not interested in sheet.				
278	M: Do you think using the materials in classroom is enough for you to learn?				
279	AntConc.				
280	SW+N: To be honest, if we can take the materials home we do not normally revise them.				
281	Instead we prefer doing the activities in class, not outside classroom.				
282	M: Do you think the materials and the progtamme help?				
283	N+M: It helps us a lot.				
284	N: I somewhat understood what you have taught us. It helps me a lot in the learning words				
285	then before.				

287 But we have **more knowledge** in LBs. We can still be progressing in LBs, not perfectly.
288 I think we need examples of LBs use in the context and we can produce something related to
289 LBs. We can do it but not perfectly.
290 M: How useful are the materials?
291 W+N: Of course, we can use it as supplementatry in English, revise and recall the LBs
292 M: How useful is the teaching of lexical bundles ?
293 W: It is very useful such as the word ' the use of', I notice how it is used in the sentences,
294 come together as a pack. I can add this LBs into the sentence. It is easier to use.
295 M: Can you remember those LBs?
296 W+N: Yes, we can remember the LB 'the use of' **easier** than using a single of such as 'use'
297 N: I think my structure has been developed.
298 W: I think if I learn a word individually, I need to memorize and say it out loud.
299 But for the LB just take a whole chunk like 'the use of', in order to'
300 M: When you study the LBs, have you found other words or LBs connected ?
301 N+W: Yes, I think the LBs are likely to follow the noun. I have to look through the
302 concordance whether the target LBs can follow the noun.
303 M: Do you think numbers of concordance lines that you have seen on AntConc screen
304 can help you learn?
305 N: yes ()()()
306 M: To what extent do you think the ERAs sessions help improve your English skills?
307 And in what aspects (e.g. words chunk, grammar, syntax, semantic)?
308 N: structure, I know what it is
309 W: writing, I think if I am able to write I need to know everything such as SVO
310 I do not know aht is subject
311 N: I think the vocabualry and the structures are important, I know the structure in writing
312 W: I think the writing comes from looking at lots of examples
313 M: Can I say that you think the sessions help you in terms of structures and writing, looking
314 at lots of examples, concordance lines (starting from lines 1-5). To sum up, you can
315 interpret the meanings from the concordance lines. You can do it without the help of the teacher.
316 W: Yes, partly but I still need to discuss with the teacher when I find it difficult, and
317 it is beyond my ability to interpret. I still need someone to help.
318 M: To what extent do you think the ERAs sessions help improve your writing in engineering?
319 W+N: Yes, I think a lot.

packing of LB 'the use of'

320	M: Do you think you can utilise LBs in writing the abstract in your project?	
321	W+N: Yes, I think I can do it. When you asked us to write the paragraph in the posttest, I have	
322	used LBs, structure our paragraphs. We also notice the use of the LBs from the concordance	samples from concordances
323	We need to look at the samples in order fo us to produce the piece of writing, especially	
324	in the specific types of writing, such as abstracts. If we just write from our previous knowledge	
325	and not looking at how other people in the field write, our writing works might not be	
326	understandable. We agree that the sessions are super helpful and have advantages. We can use	
327	this strategy in our own writing, acdemic I mean.	
328	M: Do you have any comments on the training sessions?	
329	W: I think the learning is okay, except out timetable might not fit some. The classroom	good environment
330	environment is okay, I feel like you are not a teacher, you are the facilitator who we can ask	relationship to the students
331	for help, not very control but not very loose.	facilitator rather than the teacher role of teacher

7. Jitrada&8. Natjira

M: What do you think about the training sessions? (Do you like it? How useful

J: I like it but I think the training time should be longer. We have studied only 2-3 weeks.

I have no ideas of the **word**s. If you did not explain in Thai, I could not have understood.

N: I think I have used my thinking a lot, analysing skills. I consider thses skills difficult because

my **English skills is poor**. It is very useful to use LBs in the higher level, reading skills

I think the training sessions duration is short.

M: What aspects of the training sessions do you like most (e.g. learning materials, content, etc.)?

Why?

J: I did not like the contents because it is all in English. I am not good at English. However, — english ability

you have explained what I did not understand in the classroom. Sometimes you had to explain — using L2 is hard

to me again and again in th sessions. **I enjoyed a lot** when I can answer the questions in class.

N: I think there are a lot of things in the sessions that I need to understand, and with the short — not fully controlled by facilitator

periods of time in training. I did not understand what are in the sheet, but I can ask for support

from you anytime in the class. That is a good thing.

M: Anything else?

J: To be honest, I never seen and studied the LBs before such as the AntConc, structure SVO.

I have never seen the LBs, in particular I have never looked into the words that used in the my

electrical field. I normally study only an individual word, and only general words.

N: I have seen these in the exam.

Q: Did you say you have seen these LBs and then?

N: Yes, I have seen the LBs in the exam, but I have no ideas what they are.

M: Do you think you can make sense of them after attending the sessions? Do the LBs help?

N: I think I can understand them better than before.

M: To what extent do you think the materials help?

N: Not really because there have been written in English, some I understand and some I don't.

J: First, I did not understand the concordance lines, but I did get it when you gave me examples of — examples

how to read, and interpret them. For example, you have given line 1 - 5 as a starting point first, and — concordance lines methods

then we can look through the later concordance lines. I did not get anything in the first session. — adapt to suit with the ss reading from 1-5 first

I started to understand in the second and thethird session.

J: In the first session, I did not even know whether a single word can come together as 'group

of words' and I did not pay attention to them.

N: I can understand the structure.

J: I think that the LBs are the connecting word, **something that are not important.**

N: The words that occur before or after do not get into notice.

Q: Now, you realised LBs are important in your field.

N+J: Yes, exactly.

Q: To what extent do you think the ERAs sessions help improve your English skills?

N: I think the structure, vocabulary, how to use the words, i.e. words that are understandable. — grammar (recurring thinking) belief

J: I have started to search the LBs that I want to know in the programme. I have seen that — start to search LBs

371	they are connected to the 5 LBs that you have taught us.	LBs students link to LBs taught
372	M: Anything you want to add?	
373	N: I think I know the pattern of LBs?	
374	M: Do you think the sessions help you to notice which LBs come with which words in your	
375	writing i.e. energy saving?	
376	J: It deos help because I have never written something like this before. Then, when I have written	
377	and asked you to check, and you said they are not correct. You asked me to go back and notice the	
378	patterns in the AntConc again. I have done and it helps.	
379	N: I think the programme helps me a lot, because the Ras are academic, and if I can make use	why they changed from persomal genre to their project
380	of the examples in the AntConc, it can guide me in writing my project.	in the writing last part
381	M: Can you give me the examples of LBs you have used?	
382	N: For example, I compared the use of 'in order to' and 'the effect of'. How these two words are	
383	used differently? I went back to the programme and compare loads of samples from the results of	
384	'in order to' and 'the effect of'. I compare in terms of use, words use before and after as well.	
385	J: When I want to create the sentence used the word 'analysis', I found out that this word is	words that ss find connected to the LBs given
386	linked the LBs taught in class i.e. in order to, the effect of.	claim that the LBs I have taught the ss are important and
387	M: Do you have any comments on the training sessions?	high in frequency that is why they link to other words
388	J: The time should be longer.	
389	N: the materials should have Thai.	
390	J: The materials should explain more in very detailed. Should explain academic words and put them	
391	in the materials and explain.	
392	J: We have a poor background in English.	

9. Wilasinee & 10. Sineenart

M: What do you think about the training sessions? (Do you like it? How useful is it to you?)

W: I think it was helpful because I have never learnt these LBs before. Only study 12 tenses.

S: In the first session, I did not understand what you are trying to tell me. I started to get the idea when I have studied the 2nd and the 3rd LBs. You want to emphasise the structures, which words should use with which words, followed by which words i.e. the patterns.

M: What aspects of the training sessions do you like most ? Why?

W: the contents related to engineering such as, current means กระแส, another meaning is ปจจบน. ุ ั

S: I like the programme. It guides me the use of the LBs in the sentences, clear structure. Before, I study only the tense in the books. In the sessions, I have known the academic words the writers use in the Ras. I know that 'due to the' and 'because' can be used to talk about 'reason'.

M: How useful are the materials?

S: Yes, I use the materials to do exercises.

W: I think it helps me a lot, structure, meanings. I think the structure seems to be easier for me than before. — student are more aware

M: How useful is the teaching of lexical bundles (fixed sequences of words)?

W: I think in terms of writing, only use a single word might not cover. In my opinion, the use of 3-word bundle can be used in different writing styles, in particular an academic context. I think in the future I can gain a lot from LBs. — students perceived LBs as important ways of writing; academic writing

S: I think if I use the LBs, the language looks more beautiful. — support the academic writing

M: To what extent do you think the ERAs sessions help improve your English skills?

S: I think all of them, vocabulary, structure, and words that are used in my field. — specialised words in electrical field

I do not use English in my learning, so that I think the sessions help me to know the vocabulary in my field and the use of them as well. — previous knowledge; vocabulary in the field

W: I think I can use the structure more accurately. — promote more accurate use of structure

M: How about the concordance lines? Do they help at all?

W: For me, I can learn from lots of examples in the programme, can build sentences, see the differences of structure, more importantly I can produce the sentences more accurately. — learn from examples; create the sentence (process of learning?)

M: So, how about the noticing skills? Have you used the noticing when producing the sentences? — notice

W: Yes, I did. I noticed the words that come before or after the traget LBs. It did help me a lot, specially when it comes to the production of the structure. From numbers of examples, I can figure out what is the possible structure. — words before and after; use the examples to produce their own sentences; possible sentences

W+N: It can help, we do agree.

M: Overall, do you think the training sessions help improve your ability to write articles in engineering?

W+N: Yes, we think LBs can help make our writing more acdemic. — LBs use can creat more acdemic work

M: Do you have any comments on the training sessions?

W+N: The programme should be taught continously, afraid that I will forget. — more often might provide long-term memory (suggestions for other research)

11. Adisorn & 12. Sud-auem

432	M: What do you think about the training sessions? (Do you like it? How useful			
433	is it to you?)	They agreed that they will		
434	S: I like it because I can use the LBs in my project writing now.	apply what they have learnt	like	
435	I think I can use the LBs in the producing the documents, and presenting in the meeting.	in the project writing.	can use	
436	M: Ummm			
437	S: I mean 'presenting' my work to other, I can use LBs.	She wanted to use LBs in her work.	can use	
438	A: Yes, I think it is helpful, I have known the words and their use.	He agreed that LBs are useful in	think/known words and use	
439	M: The use of words in the sentences?	extending his knowledge of words	helpful	
440	A: Yes, the writing.	and use.	writing	
441	M: What aspects of the training sessions do you like most (e.g. learning materials, content, etc.)?	I wanted to know what do 'the word		
442	Why?	and their use' mean. Does he mean the		
443	S: I like programme the most because I do not need to open the textbook. It is easier to find	use of words in sentences or what?	like	
444	something in the programme, for example if I want to study the textbooks. I can upload the	He meaned the writing.	easier	
445	cripts of the textbooks in the programe. I can search the words or structure I want to know fast	She liked the programme most.	can search	
446	and conveniently.	She can search the target words or she can		
447	M: Anything else?	upload the textbooks file if she wants to		
448	A: I also like the programme can use colour to hightligh the LBs.	study the language of it.		
449	M: How useful are the materials?	He likes the programme they ways it	like	
450	S: yes, I can revisit the contents again whenever I want.	can catch his attention by the highlight.	functions of programme	
451	M: Do you think the materials can help you learn the LBs?	In terms of the usefulness of the programme,	revisit	
452	S: yes.	she can revisit the contents later.		
453	M: How?	I asked how materials help her to learn LBs.		
454	S: some LBs have different meanings. It depends on the purpose of using them.	She said that having the purposes in mind	how your writing differs from published work. T	
455	M: How about the **examples of the use of LBs** in the ERAs you have seen from the concordance lines	seems to help learn the LBs.		
456	M: Can you make use of LBs, and can you look at the LBs in the concordance lines?		LBs and concordance lines	
457	S: yes.	She made use of the LBs by looking at		
458	M: Do the 3-word bundles are helpful?	the concordance lines.	yes.	
459	S: Yes, they are. A lot.			
460	M: What kinds of learning do you normally do in classroom?	I asked about the learning they usually do.	strategies they have	previous strategies
461	S: the words like the/of/ a, single words	They have seen the importance of single word	single word	
462	M: a single word?	such as the, of, a, an article		
463	S+A: yes yes	I wonder why only those three single words		
464	M: do you think LBs is more helpful than a single word?	so I have reassured again by emphasising		
465	S: it is. I think it makes my writing beautiful, nicer, sound nice.	a single word'.	LBs than single words	promote memory
466	M: Do you think you can remember the use of LBs in the context?	She explained that LBs are easy to remember	memory	
467	A: Yes, when I use the LBs in the context.	and LBs make their writing more nice	think	
468	S: I give you the example of the use of 3-word bundles used in the context such as	She gave example of the 3-bundle, the number		
469	the word 'the number of' can signal me the quantity of something.	of' can signal the quantity of something.		
470	M: OK. Have you ever learnt this before? You can see that this bundle is easy to learn	I asked about their previous learning they usually do.	previous learning styles	
471	the number of'	They accepted that the LBs is easy for them	LBs are easy to learn	
472	S+A: yes yes	to study		

473	M: How do you feel when you seen this bundle 'the number of' is used in the context?	I asked them to give examples when they		
474	S: The first time I have seen 'the number of', I think it should be the 'number' or 'figure'	see 'the number of' in context.	first time	existing ways of predicing
475	something like that. I did not really think that it relates to the quantity, just thought it is the number	she said she thought 'the number of'	existing knowledge	
476	together in groups like the 3-word bundles that you have introduced. I think they are useful	relate to the number or figure.		
477	M: right OK. From what you have learnt 'the number of' in the concordance lines, do you think	They have seen the use of 'the number of'	seen lot of concordance	
478	you know how to use the LBs in the context?	in context and help her figure out about	lines make her	
479	S+A: Yes, we do think that.	the number of' shows 'quantity'	generalte possibility	new knowledge
480	M: To what extent do you think the ERAs sessions help improve your English skills?	They can imprive the grammatical skills	of the use of 'the number of'	
481	S: I think for me the structure. When I write, I normally bring the words and combine.	structure'. They are more aware of the	in context	more aware
482	I don't even use the verbs, objects, or somethings that I should be used. However, after the sessions	structure when create the sentences.	think	
483	I know that when I create the sentence I should have SVO.	The use of subject verb and object in the sentence.	know	
484	M: Where did you get SVO?			
485	S: From the sentences			
486	M: Any noticing skills going on?	I asked her about the noticing skills occur.	process of thinking	
487	S: I will notice from the LBs such as 'the number of', whether I should choose 'years' to be replaced	She responded that she have a thinking processgoing on		
488	after 'the number of' and no need to add 'is' or 'if' something like that.	whether to consider meaning first and then		
489	M: But then what help you to get that, I mean the verbs?	tried to figure out the appropriate use of verb		
490	S:I think the structure.	I asked her how to figure out.		
491	M: So, which one can help you acquire the struture, the programme or what I have taught you.	She used the knwoledge from the materials and the programme	materials and programme	
492	S: I think both the programme and what you have taught in the class.		think	
493	M: Do you think you have noticing skills?	I have asked her about the noticing skills	noticing skills	
494	S: Yes.	Noticing skills for her is scanning from the		
495	M: Did you start to scan from the left to the right? Do you ususally do this?	left to the right		
496	A: yes, I have started to do 'scanning'.			
497	S: Me too.			
498	M: Do you think this kind of learning has happened or never happen to you before such as	In her previous strategies she only translated word-by-word	previous strategies	translate word-by-word
499	look at the co-occurring words, words that are likely to come together?	Before she was introduced this strategies, she did not pay attehntion to the context.		
500	S: Most of the time, I translate word-by-word. I did not look at the whole sentences, whether			
501	what do these words mean in this context.	When they use strategies, they seem to be able to feel confident in their ability.	new strategies	
502	M: It means that the sessions somehow help you to improve your English, am I right to say that?			
503	S: Yes, you are right.			
504	M: you have improved your grammar, is that correct?			
505	S+A: Yes.			
506	M: Overall, do you think the training sessions help improve your ability to write articles in engineering?			
507	Especially, in your engineering project.			
508	S: Yes, a lot.			
509	M: How? Can you give me examples?			
510	S: For exmple, writing the introduction. I normally write the introduction in Thai, and then I copy		previous ways of writing	
511	and paste at google translation. The google just did the writing to me. After the session, I have just		google translation	
512	realised that the translated writing work is not understandable, not even the human language. I cannot make		programme is useful	
513	sense of the meaning that I really want to convey. The programme is very helpful because If I don't know		can search	
514	the words or the LBs that are used in the journals, or in the articles, I can search the target words/LBs,		can apply	
515	structure, and examples of the words used in the papers. Therefore, I can apply what I have learnt in			

516	the sessions right away.	
517	M: Have you looked at the examples and used a lot of them?	examples
518	S: I think quite a lot.	think
519	M: Do the examples help you?	
520	S: yes of course such as the noticing skills. For example, I have searched the word 'solar', I have seeen	noticing skills
521	a lot of numbers and some are followed by 'is'.	
522	M: you said that the sesions provides you knowledge of the use of LBs, but also the target words that	
523	you want to know?	
524	S: Yes.	
525	A: we know how the words are used in the context.	
526	M: How about the words that are used in the engineering field?	
527	S: Yes, the words used in the electrical fields that we have never seen before.	
528	A: Yes.	
529	M: Anything you want to add more?	
530	S: I think the seesions time should be extended. And you should use this teaching in our curriculum.	
531	Normally, the teachers just taught us grammar, but did not go in-depth into the meanings.	
532	M: The teachers did not specify on the words that used in the elecrical fields because they want to emphasize	
533	to general English.	
534	S+A: yes yes	
535	M: Anything else?	
536	S+A: We think that is all.	

	13.Pimwipa	explain briefly what happen?	Code	Main themes	Sub-themes
537	M: What do you think about the training sessions? (Do you like it? How useful				
538	is it to you?)				
539	P: Yes, I liked it a lot. I have never learnt this before. I think I will use it in my future	He agreed that he will use what he has learnt	like		
540	M: Do you think the sessions helpful?	in the future.			
541	P: Yes. I never know that one word can have different meaning, and can be used with	different meaning of the words	never know		
542	different sentences.				
543	Do the seesions help you with your study?				
544	P: yes, I think I will use it in my research writing, it think it would be helpful.	he will use the knowledge of LBs in writing	think I will use		
545	M: In what aspects do the sessions help?				
546	P: I think the writing session.	He agreed that the sessions help him with	think		
547	M: How the sessions help you in writing?	his writing. In particular the single word			
548	P: For example, an individual word can be used to write in different sentences.				
549	M: What aspects of the training sessions do you like most?				
550	P: I like the programme, if I did not use it I think it is not useful.	He likes the programe	like		
551	M: Are you trying to say that the progrmme helps you a lot, in what ways	the programme helps him a lot, lots of examles			
552	such as providing lots of examples or what? Give me examples.				
553	P: The examples I have learnt from the programme are also helpful.	examples from the programme are helpful	have learnt		
554	M: How? For example, you look at lots of examples, you can apply in your writing?	I asked whether he thinks examples from the			
555	P: I mean I have seen different sentences that only use the same word.	programme can apply in writing	seen different use of LBs in sentences		
556	M: Is that it? You said you have programme and you look at the examples there with only				
557	one word used. One word can be written in different sentences.				
558	P: Yes.		yes		
559	M: How useful are the materials? I mean do they help you learn LBs?				
560	P: Yes, I joted down the things there.	materials	yes		
561	M: Do you think the materials should be improved? How?		yes		
562	P: It should include the bunches of examples, I might not bring my laptop.	I should include all examples in the materials	should		
563	M: Anything else?	because she could not follow on the screen.			
564	P: No.				
565	M: Do you think the teaching of LBs i.e. 3-word bundles is useful? How?	I asked about the usefulness of the bundles.			
566	P: Yes, I think a lot. Generally, I don't know the 3 words can come in group, and can have	She said that the bundles is new to her. In her	think/does not know		
567	meanings. I do know a single word.	previous knowledge she has not realised that the			
568	M: Do you mean the 3-word bundle differs from a single word?	words are able to come in group. She only know			
569	P: Yes, I think when I read the text 3-word bundle looks more formal.	a single word. She think that the bundles is more formal	think		
570	M: Do you like this kind of learning?	I asked her whether she likes the learning of bundles			
571	P: Yes, I like it because it is more than one word.	She liked the bundles because she think they	like		
572	M: Do the LBs make you remember easier?	look more formal and make her remember easier			
573	P: I think so. I think I can remember the LBs easier than a single word. If the word comes alone,	than the single words.	think/can remember		
574	I might not know the meaning. If I see the LBs, I feel familiar with them.	She reinserted that even though she does not	feel familiar		
575	M: Am I right to say, for example the word 'the number of' you can remember than the word	know the meanings of LBs, but she can still	familiality		
576	number'?	manage to get the meanings.			
577	P: Yes, you are. I might not remember the number because the word 'number' when it comes	She confirmed by giving example of the word	sample of LBs		
578	in a full sentence, its meaning might change.	number' that the meanings might change when	the word 'number'		
579	M: Do you think the words that co-occur with the 3-word bundle, words used before and after	used in the sentence.	and 'the number of'		
580	or the words that around the LBs can link you to new knowledge. I mean if you search the	I asked her more about the words surrounded the	LBs link to		

581	target LBs, you find out that there are a lots of words that are linked to the words in your	LBs, whether they help her in the engineering	the (specialised) engineering words	
582	engineering field?	field.		
583	P: Of course, I can remember the 3-word bundle rather than a single word, I notice from the word	She further explained that the bundles can help her to	can remember/notice	
584	that use before or after, and see whether they are nouns or what?	remember well, and can identify the word functions.	see	
585	M: Do you mean the 3-word bundle is far more easy to remember ?	I want to knwoexactly whether the 3-word bundle		
586	M: can you give me examples of the words you learnt and to what extent the LBs help improve	is easier to remember, and she agreed that.	yes	
587	P: Yes.	The LBs for her is considered helpful in writing		
588	M: Do you think it is helpful in your field?	in her workplace, especially writing and reading.	think	
589	P: Yes, I think. **I think** it can be used in writing and when I read others'work in the workplace.			
590	M: To what extent do you think the ERAs sessions help improve your English skills?	I have asked her whether the sessions help improve	think	
591	P: I think the writing, at least I have written and I still feel familiar with how the LBs is used	English skills. She felt familiar with the LBs	feel familiar	
592	in the writing. I want to use the LBs in my 'job application letter'	when she wrote. He want to apply the use of the LBs	want to	need
593	M: Do you see examples from the programme or what?	in her job application letter		
594	P: Yes, I have seen bunches of example and need to refine them again in order to write	She responded that when she see the examples from	strategies when dealing with concordance lines samples	
595	my own.	the concordance lines, she had to understand and	have seen	
596	P: Yes, of course. I should think very carefully before I do the writing.	be able to apply the samples in her own writing	need to refine	
597	M: In what ways, do the examples help you?	I further asked her about how examples help her	should think carefully	
598	P: Yes, I can see the ways how the LBs are used differently in sentences. The first sentence I	in her writing.She explained that she can see	can see	
599	can notice there might be the use of Subject and Verb, the second one might have the use of	different use of LBs in sentences, focusing on	can notice	
600	adjective. I have seen lots of sentences.	the grammatical use.	have seen	
601	M: Any noticing skills?			
602	P: Yes.	I asked her did she use the noticing skills.	noticing skills	
603	M: do you think the training sessions help improve your ability to write articles in engineering?	I asked her whether the sessions help improve		
604	P: Yes, it was difficult first, but now I understand.	her writing especially the articles writing.	I understand	
605	M: I mean the sessions help improve your ability to write articles in engineering.	She accepted that it was hard in the first time		
606	P: I think I will change the texts to the txt. File, and look at the examples. I will use my articles	but now she understood.	change the txt. File	apply for her own work
607	that are translated in English and look whether my sentences are correct.	She added that she can even use AntConc to look	look at the examples	
608	M: Do you have any comments on the training sessions?	at her writing whether it is correct by comparing to		
609	P: I think I want you to open a new course that brings the LBs in training. I don't like the current	other people work.	want	
610	English courses provided. All English courses are only conducted like remember all the	She has suggested that the LBs should be used as	do not like	
611	scripts in the conversation and do the role play. I think I will rarely make use of that teaching.	contents in English courses. She did not like	the current English	
612	Therefore, the LBs teaching should be promoted in our uni.	the convention of the current English courses.	rarely make use	

Appendix (11). Sample of a field note

Tuesday 16, February 2016

Main activities:

- Pre-Test
- Introduction to the lexical bundle's workshop

I distributed the pre-test to 24 students. While they were working on the pre-test. I noticed from their facial expressions that they had difficulties working on the pre-test. After they have finished the test, they confessed that they have never seen the lexical bundle before. particularly in part 2. They misunderstood the question in part 2, they understand that they have to mix up of the topics. My fault, I think. I should have made the questions clearly for them.

Friday 19, February 2016

Main activities: Session 1: Introduction to the concept of lexical bundles in ERAs with the aid from AntConc. In this session, I have introduced the concept of corpora to the engineering students. Students were working on the activities of the provided materials. In this way, the students were introduced AntConc and working with AntConc activities. The students enjoyed playing with AntConc, and most of them could follow the steps in uploading the software and the 25 ERAs txt. files. In order for the students to get started to work with the lexical bundles, I am familiar them by introducing the lexical word, i.e.,house. By doing so, the students went through the questions in the activities. Then, the students went through the lexical bundle's activities. For example, the students search the bundle 'the number of', and come up with the overall meaning of the target word. Then, they went through 5 concordance lines until they can identify the meanings of the target word in the ERAs corpus.

I had planned that if the students cannot pick up the meanings and the structure of the bundles, they need to interpret until line 202. Luckily, they can identify the meanings and structures of the target bundles when they reach line 20. I remind them to make sure the meanings they come up was the meanings that the 202 concordances represent. While the students were working on the concordances, I was there to facilitate them when they need. I notice that some of them did not really follow the steps so I had to explain to them. Or some of them ask their friend to help. Overall. I think they enjoyed the session even though they could not fully understand all of the words that are also surrounded. The words that are used with the target bundle 'the number of' are also taught. Then, facilitator (me) shared the answers of the activities with the students.

In terms of the time, the class takes more than 2.5 hours since the class moves very slow. I think it is normal because the students just had a chance to expose the use of the ERAs corpus. I am very satisfied with the students' engagement even though it is very challenging for them in doing the activities. Moreover, I think the students felt at ease since the classroom environment is very friendly. They can ask for help anytime they

want, and I also did the activities at the same time with them. I kind of share the answer with the students. Interestingly, I think the students' anxiety decrease at the very end of the first session. Unlike the first session, I could notice that the students were nervous and were not sure about the workshop. I think the students needed to expose more activities and practice interpret and make a decision about the meanings of the target bundles. For the time being, I need to guide them a lot about what to look for and how to interpret and pinpoint the meanings of the target bundle.

Appendix (12). Examples of students' activities and students' notes

Table 1

Item	groups of words	Words that co-occur	Types of words	Indicate what (แสดงถึงอะไร)
1.	the number of	committed	noun	จำนวนที่ทำ
2.	The effect of	percentage	noun.	การแสดงผล.
3.	the use of	a high end computer ~~a multi physic~~	noun.	การใช้คอมพิวเตอร์
4.	The use of	a multi physic	noun;	การออกแบบระบบต่างๆ
5.	The use of	a networking cloud	noun.	ออกแบบการสร้างระบบอินเทอร์เน็ต
6.	The effect of	DG units.	noun.	การศึกษาโมเดลโดยต่างๆ.
7.	the effect of	different	adj	ความแตกต่างของอุณหภูมิ
8.	the numbers of	buffers.	noun.	ตัวช่วย

due to the

status of material			
existence and non-existence	action	Quality	other
deterministic	lack	buffer	amount
absence	action	increased	complexities
combinatorial	charging	low	criticality
limit	acceleration	nonlinear	electric
	additional	better	fundamentally

1. เป็นการเรียนรู้คำศัพท์ที่เป็นกลุ่มคำ หรือที่เรียกว่า Lexical bundles กลุ่มคำที่ได้ศึกษามี ดังนี้
   The number of , The use of , The effect of , the presence of ,
   due to the , in order to

2. การใช้ The number of จะตามหลังด้วยคำนาม
   The use of ตามหลังด้วยคำนาม
   The effect of ตามหลังด้วยคำนาม
   The presence of ตามหลังด้วย
   due to the ตามหลังด้วยกลุ่มคำนาม
   in order to ตามหลังด้วยกลุ่มคำนาม

3. ได้เรียนรู้ถึงความหมายของคำ

4. ได้เรียนรู้การใช้ Patterns ของแต่ละคำ เช่น
   The number of : the number of + plural noun + singular verb
   The effet of : V + the effect of + n
   The presence of : n + the presence of + adj
   The use of : the use of + n + v + adv
   due to the : due to the + Noun Phrase , effect + due to the + clause
   in order to. : In order to + NP ... , S c by ing) + O passive form
   S + is to + V, of + comprement + in order to + V, + NP and NP

5. ได้เรียนรู้คำศัพท์ทางวิศวกรรม ใน concordance line และสามารถบอกสถานะของ materials ได้
   เช่น absence จะบอกความเป็น Existance and non-existence
   acceleration บอก Action

6. ได้เรียนรู้การใช้โปรแกรม Antconc
   ปัจจุบันได้ศึกษาและเรียนรู้การใช้โปรแกรม Antconc ทำให้การเรียนรู้ภาษาอังกฤษได้เข้าใจมากขึ้น และในอนาคต การใช้โปรแกรม Antconc จะเป็นประโยชน์ และสามารถนำมาปรับใช้ในการเขียนทักษะการเขียนภาษาอังกฤษของตัวเองได้มากขึ้น

Appendix (13). Materials that are used in the classroom

Lexical bundles in engineering research and AntConc concordances workshop

In this workshop, there are two main sessions: knowing corpus and concordance, and using AntConc concordance to examine the data in a small corpus of the engineering research articles. The second session introduced concepts of the lexical bundles. By the end of this lesson, you will be able to

know what a 'corpus' is?

learn how to work with concordance and use AntConc concordance in the workshop

learn how to interpret the concordance lines

know forms, meaning, and use of the lexical bundles in Engineering Research Articles (ERAs)

Part 1**:** Getting to know ERAs corpus and concordance lines

Have you ever heard about corpus? What is a corpus?

A corpus (plural: corpora) is a collection of electronic texts. Corpora are built with a specific purpose in mind and are designed according to specific criteria. For example, if you want to check the language used in your field, you can build a corpus of relevant research articles written by experts. Corpora are accessed by using text analysis software. (Hunston, 2002; Charles, 2009)

In the workshop, we will use a corpus named 'Engineering Research Articles Corpus' (ERAs corpus), which is compiled to help learn the use of lexical bundles from the engineering research articles.

Activity 1:

Focus: concordance lines

Purpose: to get to know the concordance lines

Instructions: Have you ever heard about corpus? Have you ever used concordance lines before? If not, scan the following lines 1 to lines 10 of the text taken from British Academic Written English Corpus **(**BAWE**).**

Focus on the word 'house', and try to answer the questions that follow. You can discuss with the person near you. The answer will be explained and discussed together as a whole class.

N	concordance			
1	what was 'really going on' at the halfway	house	- what was 'really going on' was residents
2	'one-stop shop' for advice on benefits,	housing	, education, lifestyle, etc. They also provide
3	nationality, education, economics, distribution,	housing	, labour protection and social security
4	classic' welfare state: health, education,	housing	, and the national insurance and national
5	responsibilities, which include: education,	housing	, planning applications, strategic planning
6	level 24. This includes improving education,	housing	, material circumstances, targeting young
7	would be used for investments in education,	housing	, health care, or job training. Contrary
8	inequalities in many areas, including education,	housing	, and employment. Ethnic minorities are
9	that I expect to look at how the feeds,	housing	, fertiliser, etc. can be improved to meet
10	and environmental factors such as income,	housing	, education, safety, access to fresh food

Table 1*:* concordance lines 1 -10 taken from BAWE

Questions:

1. What is the meaning of the word 'housing' in this context??
2. What are the types of words used before ‘housing’?
3. What type of word is most frequently used after ‘housing’? What are the other types of words used after ‘housing’?
4. What are the patterns of ‘housing’ used in the text?
5. In general, what is the meaning of the word house in the above concordances?

Note

_____

_____

_____

The possible answer of activity 1 is on the next page.

Possible answer:

1. The word ‘house’ or we call it as the ‘node’ word or the ‘target’ word.
2. Nouns
3. There are individual words or groups of words.
4. The house is frequently used as object in the sentences. The possible pattern (Line 6) for example, subject (This), Verbs (includes), Object (improving education, housing…)
5. The word ‘house’ means the normal house (in line 1), as you can see from the words that associated with the house e.g.,halfway, residents. Other meaning of the house referred to one of the government benefits systems (can see from line 2 to line 10). It can be noticed that the target word ‘housing’ used in the written words here seemed to be formal.

Activity 2**:**

Instructions: Notice the use of the word ‘house’ taken from British Academic Written English Corpus (BAWE) from Activity 1.

Do you think the use of the word ‘house’ from Activity 1 differs from the use of the word ‘house’ in British Academic Spoken English Corpus (BASE), from Activity 2 ?

_____

_____

_____

Table 2*:* concordance lines 1-10 taken from BASE

N concordance			
1	nineteenth century a novel called The Counting	House	[[voiced pause]] and by the way i use the
2	[[voiced pause]] in the [[voiced pause]] great	house	and then a younger servant a younger girl
3	it was within the compound of the great	house	surrounded by a high wall of Suffolk brick
4	Gladstone graveyard Anglican church plantation	house	my grandpa work so hard that he learn Latin
5	wood from the fence surrounding the great	house	or tools from the warehouse with the money
6	chapter from [[voiced pause]] this Counting	House	book right by the way i should say as a
7	[[voiced pause]] displaying them at in his	house	in London and allowing paying visitors
8	series [[voiced pause]] of discussions in the	House	of Commons through eighteen-fifteen and
9	estates with their rich and well equipped	houses	in the country and which was the worst
10	things like the woman being the angel in the	house	well most women in the nineteenth century

Now, you should know what concordance line is. Can you write the answer on your own word? Or you can write the answer in Thai. Then, we can share our answer on the board together. Then, check the meaning of concordance on the next page.

_____

_____

_____

Possible answer of Activity 2: The target word 'house' was used in the BASE (from lines 1-10) means the normal house. You can see from the words that associated with the house e.g.,well-equipped, angel in the, his. As shown from lines 1-10, the word 'house' was normally used in the spoken language and it was more informal from the word 'housing' from Activity 1 above.

Possible Answer:

Concordance is a list of lines of text containing a node word, nowadays generated by computer as the principal output of a search of a corpus showing the word in its contexts and thus representing a sum of its usage. Also, we usually study a few words to the left and right position of the search term" (Teubert and Cermakova, 2007, p. 104; Baker, 2007: p. 71).

You will learn how to sort the target words in the next section (see Part 2, Screen Shot 3: Concordance on the word 'current').

Note

_____

_____

Part 2**:** Introduction to AntConc

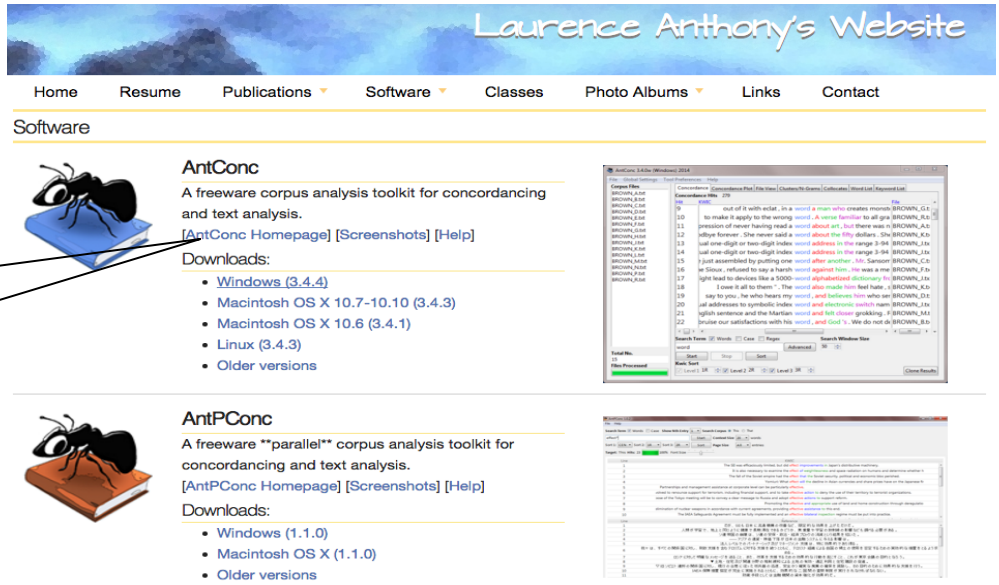
Focus: Using AntConc to explore a small corpus of engineering research articles ***(***ERAs corpus***)***

A guide to the use of AntConc in the sessions

We will use the AntConc software, version 3.4.4 and you can freely download from http://www.laurenceanthony.net/software.html. There are versions for Wondows, Mac and Linux. Further information is available on the website.

1. Start up Internet browser and go to http://www.laurenceanthony.net/software.html. you will see the following page.

Screen Shot 1: Loading AntConc from the website



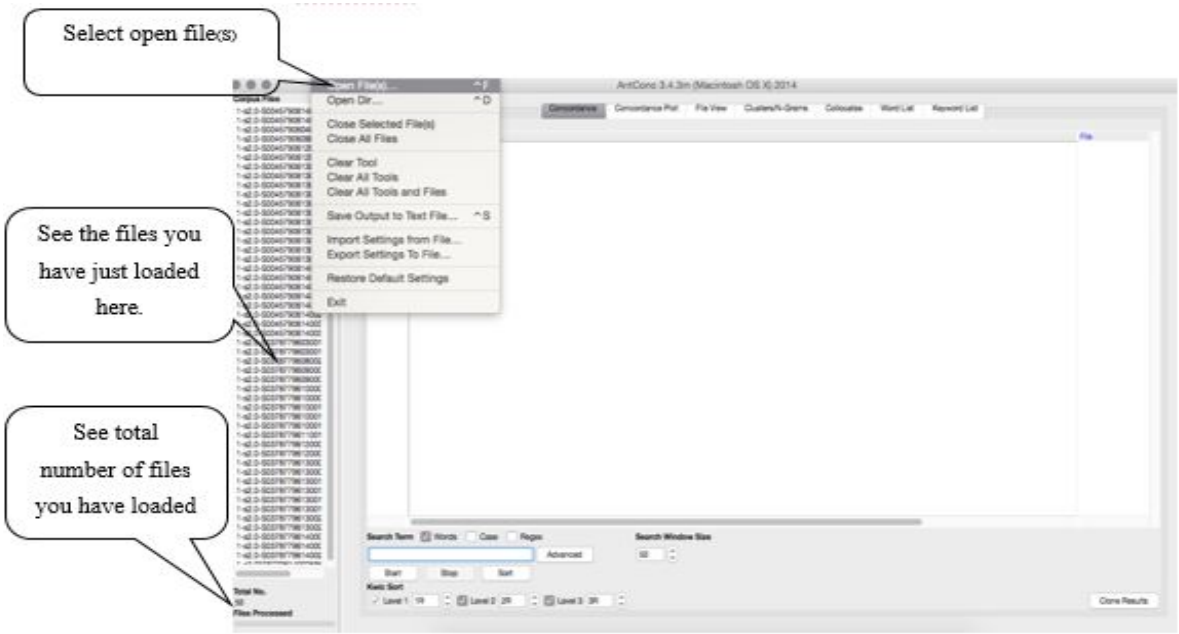
Click on Window version 3.4.4

2. Click on Window version 3.4.4, and you will be able to run the software. Now, you will be able to use the software.

3. On the file menu, select open file(s).

4. Select the folder where you have put your corpus and click OK. You should be able to have 50 txt. files (from computer and electrical txt. files).

Screen Shot 2*:* Loaded AntConc with a Corpus

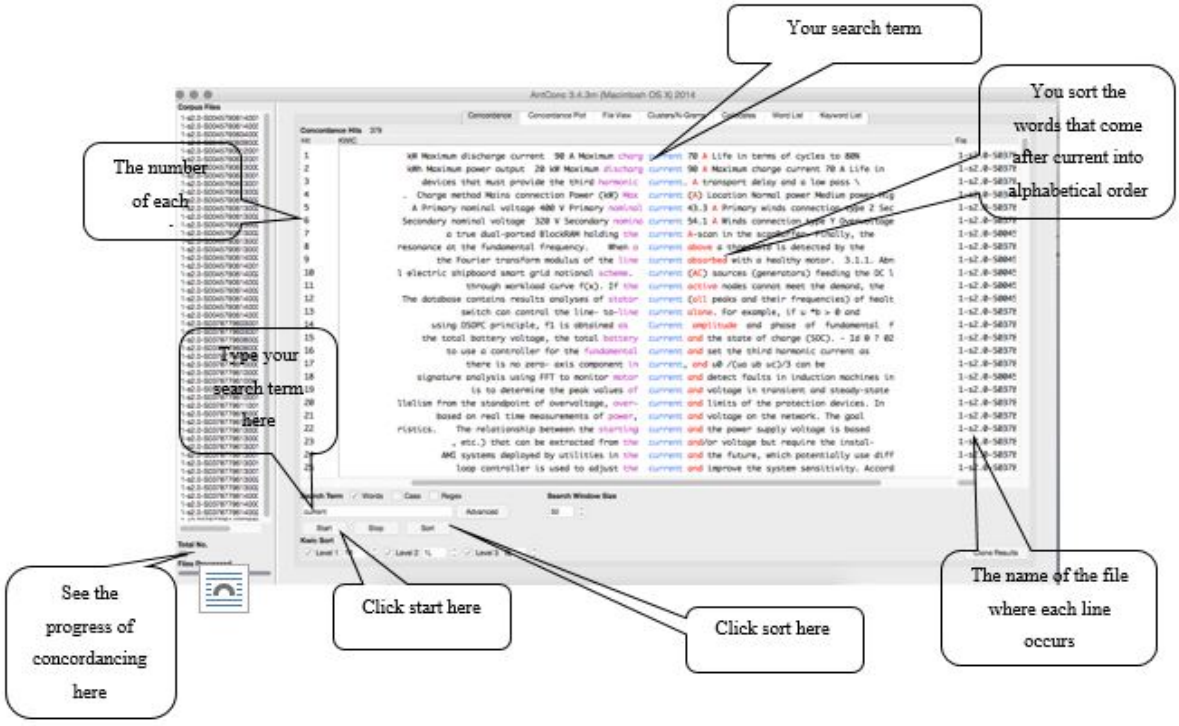


5. Type the word or phrase you want to find in the Search Term box. Type the word 'current'.

6. Click the *start* button below the box.

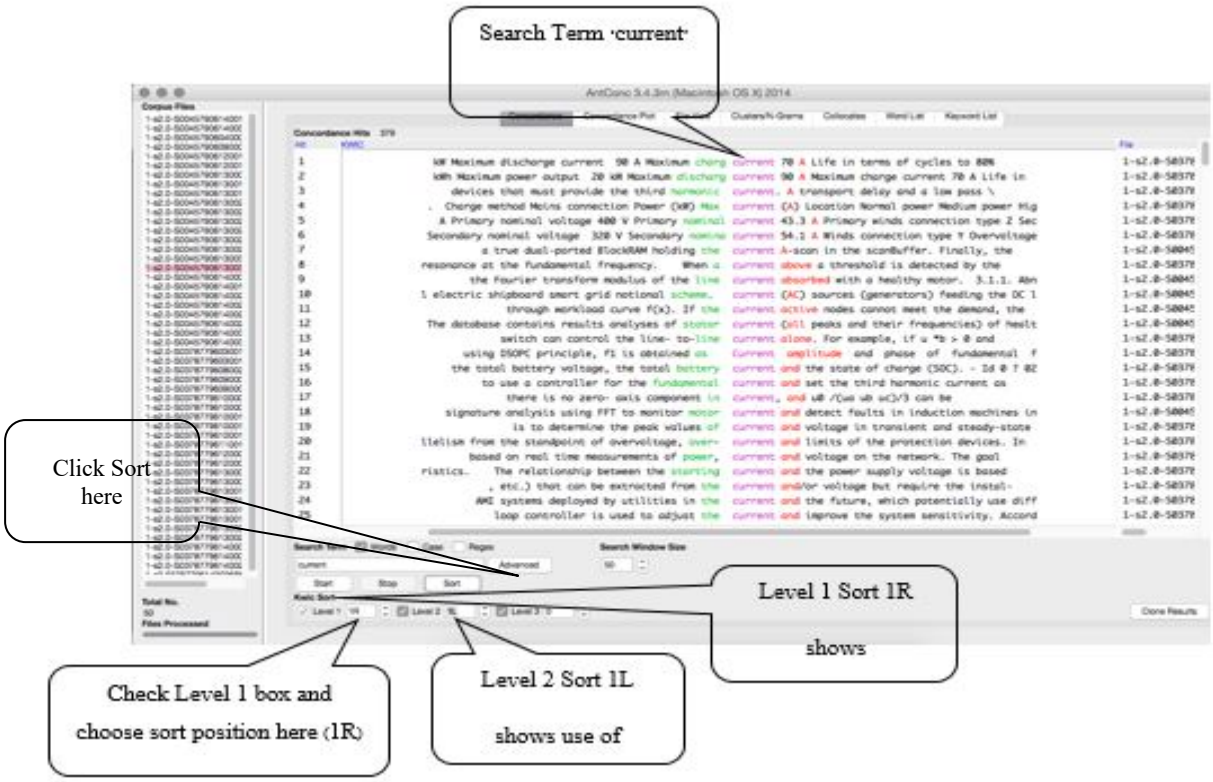
7. The concordance lines appear in the main window, with your search term highlighted in the centre.

Screen Shot 3*:* Concordance on the Word *'current'*



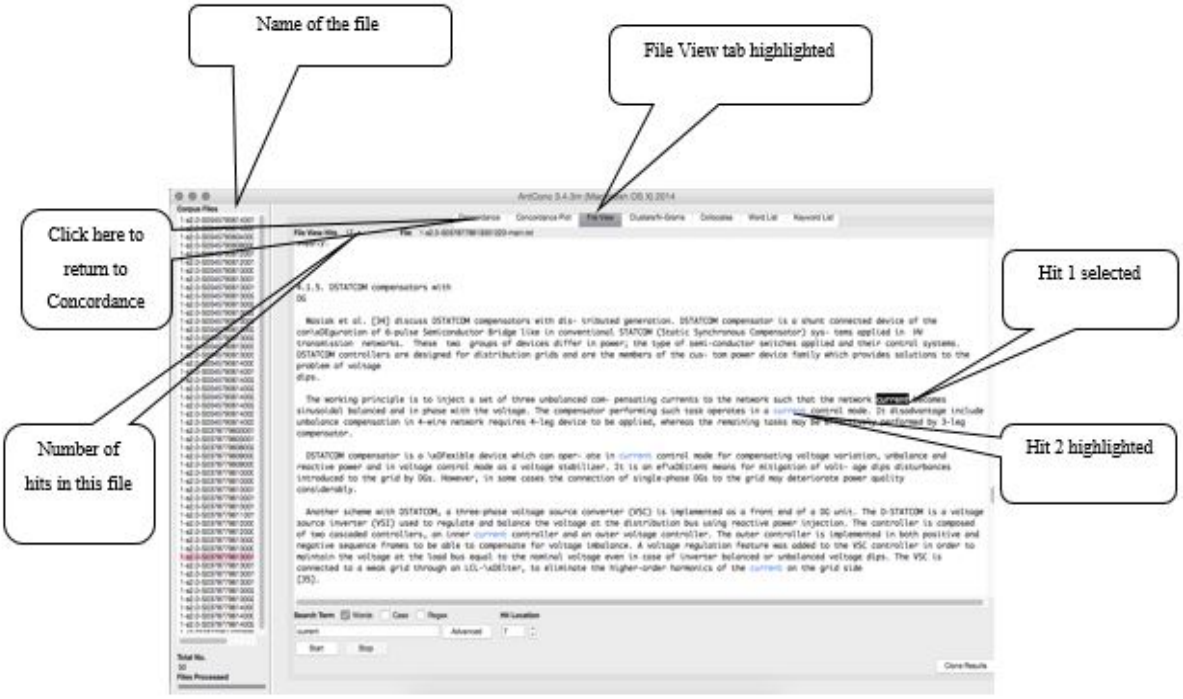
8. Click *sort* and you can sort the words that come before of after the target word into alphabetical order (from a to z).

Screen Shot 4*:* Concordance on the Word *'current'* Sorted by Level 1*:* 1R and Level 2*:* 1L



9. Change the basis of the concordance sort for example, 1R means one-word-to-the-right of *current,* 1L means one-word-to-the-left of *current,* and so on. For now, we want the 1R and the press *sort*.

Screen Shot 5*:* File view of the word *'current'*



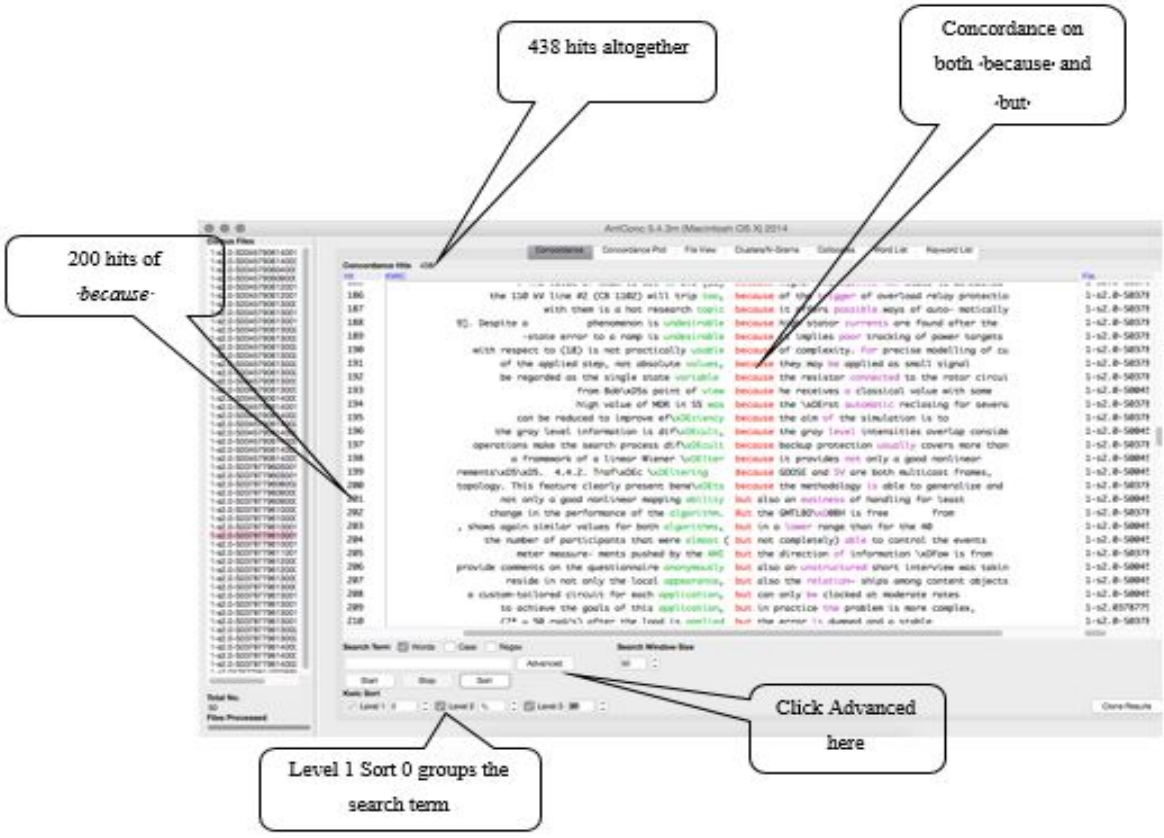
10. Click the search term, you will see the original file with the term '*current*' selected. It will bring you to the File View highlighted at the top of the window.

11. Look at the top of the window, it gives the total number of occurrences of your search term within that file and the name of the file. Here, the total number of the word *'current'* occurs in this file is 17 (hits).

12. Click the Concordance tab at the top of the window order to return to the concordance.

Screen Shot 6*:* Concordance on the Words *'because'* and *'but'*

at the same time



Screen Shot 7*:* Advanced Search Window for Search Terms *'*because*'* and *'*but*'*

13. Click Advanced Search to search for two or more terms at the same time.

14. Tick on the Use search term(s) from list below. Here, want to search the terms '*because*' and '*but* 'at the same time.

Session 2.1**:** Lexical bundles forms in engineering research articles

Activity 1**:** where are the lexical bundles in the sentences?

Instructions:

1. From the nine sentences below, underline *the use of, the effect of, the number of*.

2. From the provided sentences, notice how each lexical bundle (the use of, the effect of, the number of) is placed and used in the sentences, such as in the beginning or the middle of the sentence, as a subject, or as an object.

1. The use of a local product depends on its specific properties and the requirements for a particular job.

2. If their claim is valid, then adequate protection of concrete should be ensured by the use of a low w/c alone or by the use of Type V cement alone.

3. The integration has been possible thanks to the use of a customized communication protocol, whose details of implementation are discussed. Finally Section 5 reports the conclusions.

4. It is instructive to see the effect of additive on creep and relaxation behavior.

5. The effect of Ca2+ intrusion reflects itself in the different gradings of SAP D.

6. A subroutine was written to consider the effect of conduction, convection and radiation during the laser deposition process

7. We analyse the influence of (a, b) pairs on MSA features as well as a range of suitable values, while pointing out the relationship between the number of (a, b) pairs and the recognition rate.

8. The number of added FFs should be large enough so that there is a high probability that PUF response sets the initial power-up state to one of the added states.

9. Where N is the incremental frequency (the number of AE hits with amplitude greater than the threshold), M is earthquake magnitude or Richter magnitude of events, a is an empirical constant and b is the AE based b-value.

If you are not sure what the subject or the object is, you can go to *self-study section* below in order to remind yourself of the subject and the object used in the sentences.

Self-study section**:** Subject and Object

Subject

The subject is a noun or a pronoun that comes before verb in an ordinary affirmative sentence. An affirmative sentence is often called active sentence or who or what does the action that the verb refers to (Adapted from Swan, 1996).

For example, *the engineer* is doing the experiment. This sentence is called active sentence, and the engineer is the subject and he does the action (experiment).

Object

The object can be also a noun or pronoun that normally comes after the verb, in an active clause. The direct object refers to a person or thing affected by the action of the verb, for example Take *the dog* for a walk (The dog is a direct object of this sentence). The indirect object usually refers to a person who receives the direct object e.g.,Ann gave *me a watch* (me is an indirect object and the direct object is a watch) (Adapted from Swan, 1996).

Activity 2**:**

Instructions: On your own, look at the words in bold whether they are the subject (S) or the object (O). Compare the answer with your friends.

1. Radial basis function (RBFs) neural networks as a kind of powerful kernel methods have been applied to many areas with success.
2. The theoretical analysis of RBF structures and algorithms includes the orthogonal least square algorithm, the approximation capability analysis [17,18,20], the design of RBF structure using fuzzy clustering method, the optimization of RBF structure using kernel orthonormalization method or combined supervised and unsupervised learning method, and the use of Fisher separability ratio for the selection of RBF centres.
3. The RBF is selected because of its compact support [17,18,20].
4. Regarding Wiener, it has been implemented with different structures for noise suppression applications as MEG signals, acoustic signals, etc. [21].
5. The effect of percentage of VPA on the permeability of VPC at the age of 28 days is shown in Fig. 2.

Now, you know what the subject and the object in the sentences. So, try to review them again by doing Activity 3 below. You can discuss with your friends near you, if necessary.

Activity 3**:**

Instructions: Can you underline the subject and the object of the following sentences.

1. The permeability is increased from approximately 3.6 x 10 - 10 cm/s to around 13 x 10 - 10 cm/s when VPA content is increased from 0% to 100% by volume.
2. Compared to normal (0% VPA) concrete, the permeability of VPC with 100% VPA is about 3.5 times greater.
3. The changes in the number of equilibria now take place at higher values of the load.
4. The poorly damped oscillatory behaviour has been avoided producing a smoother operation.
5. In Eq. (9) the amount of lubricant delivered by the lubricator is calculated by counting the number of lubricant drops delivered in a given time period (usually 120 s).

What have you noticed from the above activities? We can say that the subject can be a single noun and can include all modifiers that go with it. And the object can be noticed by _____ _____.

Homework**!**

Instructions: 1. Study the group of words 'the effect of' from the concordance lines.

2. Notice the words that use *before* or *after* the word 'the effect of' and try to list them in the same groups.

3. From the concordance lines, try to identify types of words (i.e.,nouns, verbs, adjectives, adverbs), and patterns of those words.

4. Share the possible answer with your instructor and your friends next time.

Activity B**:** Knowing the patterns of the lexical bundles

Instructions:

1. The following patterns tell us about how the lexical bundles (*the effect of, the number of*) are used in the research articles. Can you group the right concordances with the appropriate pattern?

2. When finished, you can share the answer with your neighbours and then you will be given the possible answer together as a whole class.

3. Can you guess which section of the research article each sentence comes from?

1. However, because the quantity of chloride ions absorbed by C–S–H gel is far lower than that bound by FS [42], *the effect of* decomposition of C–S–H gel on the stability of bound chlorides is insignificant when compared with the decomposition of FS under MgSO4 attack.

2. However, the values of radial stress at the inner and outer boundaries of the cylinder are not zero because of the effect of prescribed pressure loadings.

3. When we add a large number of new states |S0|, |S| + |S0| states can be implemented by a linear growth *in the number of* FFs that is log (|S| + |S0|).

4. Actually, when the dimensions of the generator are reduced, the induced voltage is decrease inherently as a result of decrease in the total flux even if *the number of* conductor remains the same.

5. Müllauer et al. investigated the effect of external sulfate attack on the leaching behavior of heavy metals in concrete.

6. Fig. 6 shows *the effect of* pile spacing on the thermally-induced mechanical behavior of energy pile (3*3 arrangement) in a group in sand.

7. This study has enabled investigating *the effect of* the water distribution system on the thermal performance of a forced draft counter-flow cooling tower (FDCT) filled with six different types of drift eliminators.

8. The only difference is that they were immersed in a similar covered container of 5% MgSO4 solution for 28, 56 and 90 days at the standard curing condition (20 ± 2 °C and 95% RH) for the examination of *the effect of* MgSO4 attack on the stability of bound chlorides.

9. Hajidavalloo et al. [18], in their work of cross flow cooling towers in variable wet-bulb temperature, include a brief discussion about *the effect of* the drift eliminator on tower performance, taking only a reduction of airflow rate into account.

10. The parameter is attributed in such a way that the total number of generated clones is twice *the number of* B cells in the population, i.e., nc = 200.

11. The effect of frequency and power was characterised on the depth and efficiency of concrete removal.

12. A reduction in *the number of* buffers improves the power efficiency; whereas, keeping a few necessary buffers intact, improves the performance.

13. In order to evaluate *the effect of* thermoelectric leg geometries on the power generation performance of thermoelectric modules, finite-element thermoelectric analyses were performed on rectangular-, trapezoidal-, octagonal-prism, and cylindrical models for the temperature gradient of 100 ºC.

14. *The number of* added FFs should be large enough so that there is a high probability that PUF response sets the initial power-up state to one of the added states.

15. *The number of* nodes can be reduced to improve efficiency because the aim of the simulation is to explore the relationship between the TF variations and the severity of the inter-turn fault rather than build a detailed numerical model for a specified winding.

Try doing the following exercise, assigning each of the sentences above to one of these patterns. BUT, if you are not sure what they are, there is an explanatory section about *clause* and *phrase* in the *self-study section* provided for you.

Pattern 1: beginning of a clause or phrase + *the effect of* or *the number of* as a dependent complement clause or prepositional

1.
2.
3.
4.

Pattern 2: beginning of a clause with *the effect of* or *the number of* embedded + a prepositional phrase

1.

2.

3.

4.

5.

6.

Pattern 3: begins with *the effect of* clause + a prepositional phrase

1.

Pattern 4: begins with a phrase embedded with *the effect of* or *the number of* clause plus a dependent complement clause or a prepositional phrase

1.

2.

3.

4.

Self-study section**:** Reviewing phrase and clause

A *phrase* consists of two or more words that function together as a group, such as a large long machine, in this study, the effect of percentage of VPA on the permeability of VPC at the age of 28 days.

In contrast, a *clause* contains a subject and verb, and it can convey a complete idea. The clause is joined to the rest of the sentence by a conjunction e.g.,*Mary said* that *she was tired*. There are two types of clause: an independent clause and dependent clause. An independent clause can express a complete thought (and can be a standalone sentence). A dependent clause is usually a supporting part of a sentence, and it cannot stand by itself as a meaningful proposition (idea).

The clause is sometimes contained participles or infinitives (with no subject or conjunction), such as *Not knowing what to do*, I telephoned Robin; I persuaded her *to try a new method.* You can see that the first sentence, '*Not knowing what to do*, I telephoned Robin.' contains present participle ending with '-ing' form (Not knowing what to do). The infinitive 'to + base verb' in the second one is 'to try a new method'. (Adapted from Swan, 1996)

Activity 1**:**

Instructions: On your own, identify independent and dependent clauses in the following sentences. Check the answer together in class.

1. Before beginning each rheological measurement, the mortar was agitated using a spoon in order to reduce the effect of possible sedimentation.

2. Thirdly, the use of the term "Na2O-equivalent" implies that the effect of potassium and sodium is equivalent.

3. However, there are numerous studies indicating that the influence of potassium and sodium on silica and quartz dissolution is not identical [e.g.,[18,19]].

4. Despite the work on the interaction of sulfate and chloride in concrete mentioned above, there is very little work on the effect of sulfate attack on the stability of bound chlorides in concrete.
5. In order to better capture the effect of the segment on the efficiency, the TE calculation is carried out at a constant load resistance condition.

Activity C**:** Knowing about engineering actions by the use of lexical bundles
Instructions:
1. Notice the use of the lexical bundles (*the use of, the effect of, the number of*) in the following concordance lines.
2. Underline the words that follow the lexical bundles i.e., *the effect of, the use of, and the number of*. Identify what types of word usually follow them such as, nouns, verbs, collocations, adverbs. Then, complete Table 1 below.
3. Do you think the words that follow the three groups of words play an important role in the sentence, and indicate specific actions in engineering? How?

1. The risk may then be compared with a standard defined by the utility or by a regulatory authority, in order to check whether or not it is necessary to increase *the number of* committed units in that situation.
2. *The effect of* percentage of VPA on the permeability of VPC at the age of 28 days is shown in Fig. 2.
3. However, *the use of* a high end computer and the wired installation per home increases the expense of the system.
4. *The use of* a multi-physic software plays a fundamental role for the design of power systems, due to that great part of the challenges are not limited to electrical and communication network but also involve other domain and energy networks (e.g.,heating and thermal) [7]
5. Ref. [12] involves *the use of* a networking cloud, Pachube, which is a real time data infrastructure that allows the management of data points from individuals, organizations and companies through the Internet.
6. The voltage divider model was used to study *the effect of* DG units on voltage dip.
7. Although there are detailed differences in the temperature distributions because of *the effect of* different convective cooling and the different disc geometry, overall the results demonstrated that the scaling methodology can be used with confidence for the design and development of automotive disc brake systems.
8. A reduction in the number of buffers improves the power efficiency; whereas, keeping a few necessary buffers intact, improves the performance.
9. This is because *the number of* alternative paths that a header can take to progress changes as it advances towards the destination.
10. If *the number of* faulty links surpasses the available spare links, then the split transmission and packet re-organization are performed.
Table 5 Answer for Activity C

Item	Groups of words	Words that co-occur	Types of words	Indicate what
1.				

2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				

Appendix (14). Some activities that were developed after the workshop can be used as guidelines in integrating corpus and lexical bundles in teaching writing research articles in the future.

Part 1: Preparing the students for using corpus

Activity 1: Introducing A Part 1: Preparing the students for using corpus

Activity 1: Introducing AntConc

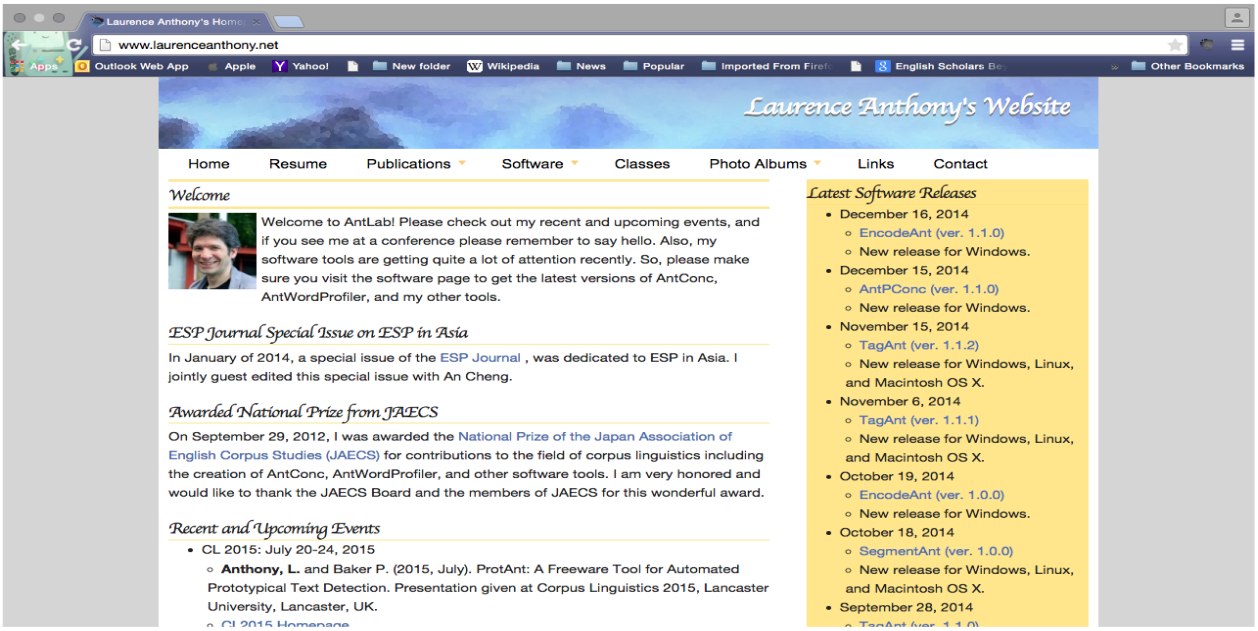
Purpose: Be able to use AntConc in the session, a guide of using AntConc is provided as follows

A guide to the use of AntConc in the training sessions

AntConc is a very useful tool and easy-to-use tool for carrying out corpus linguistics research and data-driven learning. It runs on any computer running Microsoft Windows (tested on Win 98/Me/2000/NT, XP, Vista, Win 7), Macintosh OS X (tested on 10.4.x, 10.5.x, 10.6.x), and Linux (tested on Ubuntu 10, Linux Mint). AntConc has been used by corpus linguists, language teachers, and students. If you want to know how to use AntConc, you can watch AntConc video tutorial from YouTube. In this session, you will be provided corpus data and guided on how to use the 'concordance tool' in conjunction with the training.

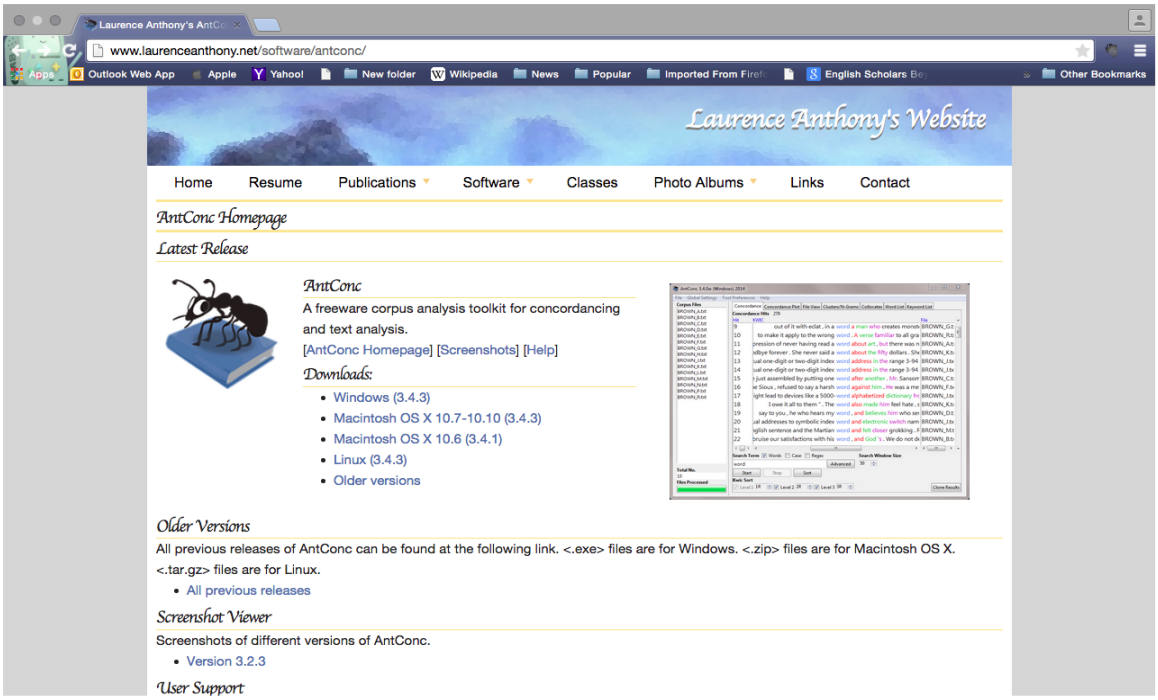
Starting AntConc software

Start up an Internet browser and go to http://www.laurenceanthony.net/. You will see the following page.

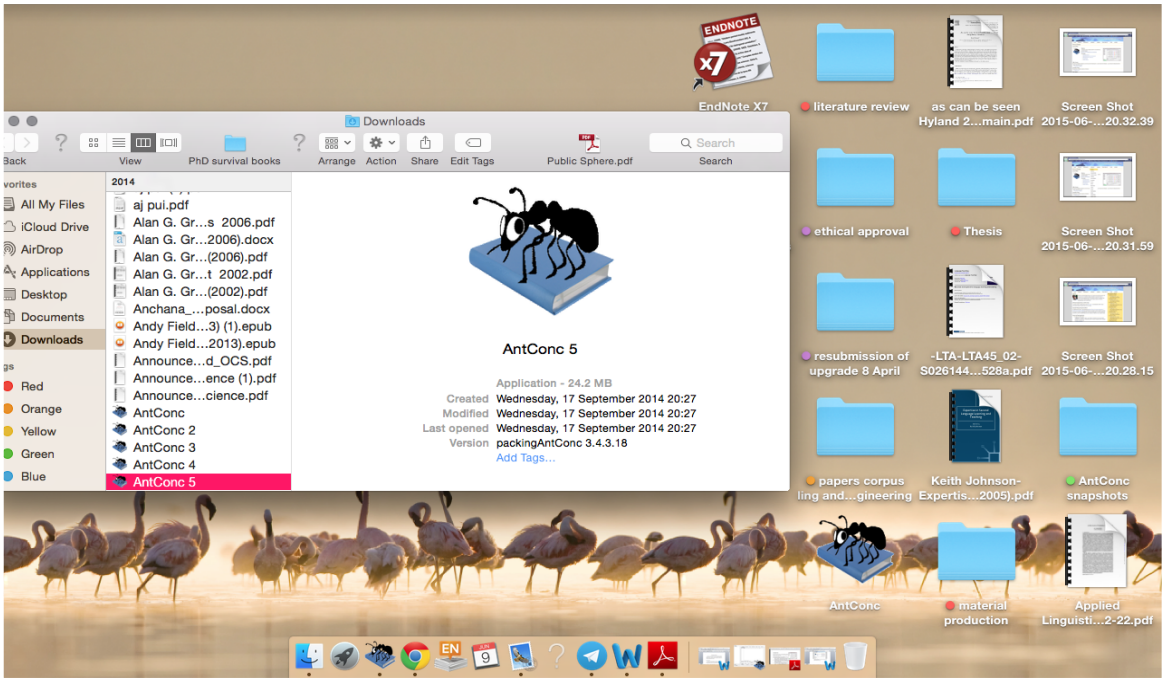


Downloading AntConc

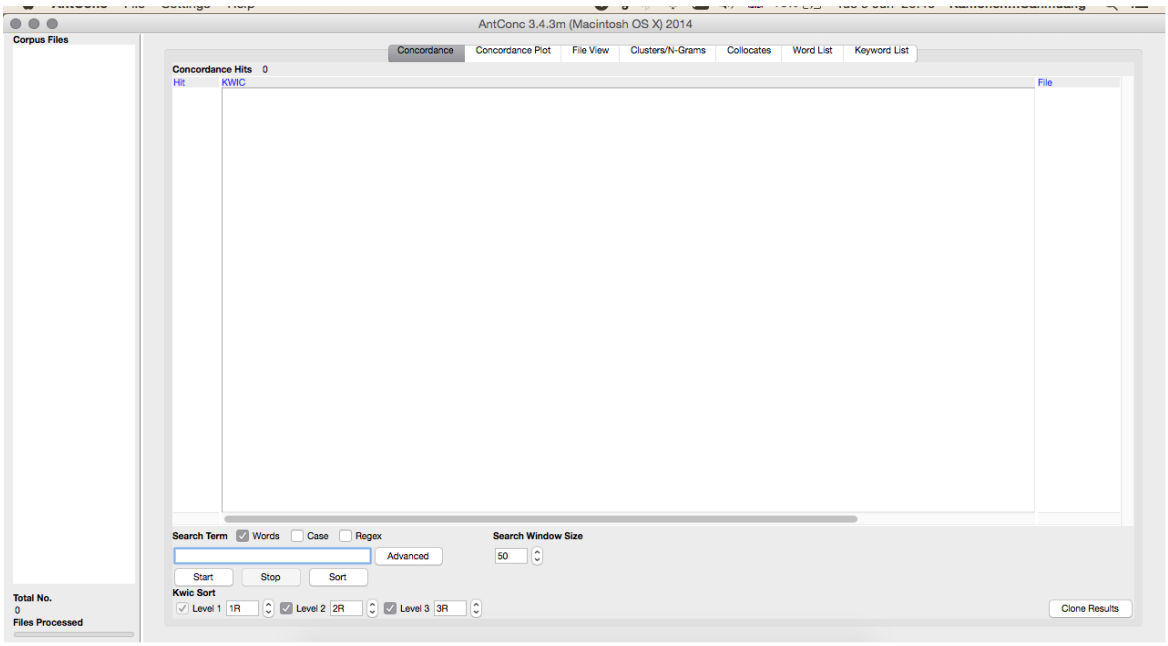
To download, click on Software and choose AntConc on the top and you will see the following page.



Click on the version that you want to download, and you will see AntConc pops up AntConc as shown below.



AntConc will appear on your desktop. You do not need to install it. Double click on the software and you can start the AntConc as shown below.

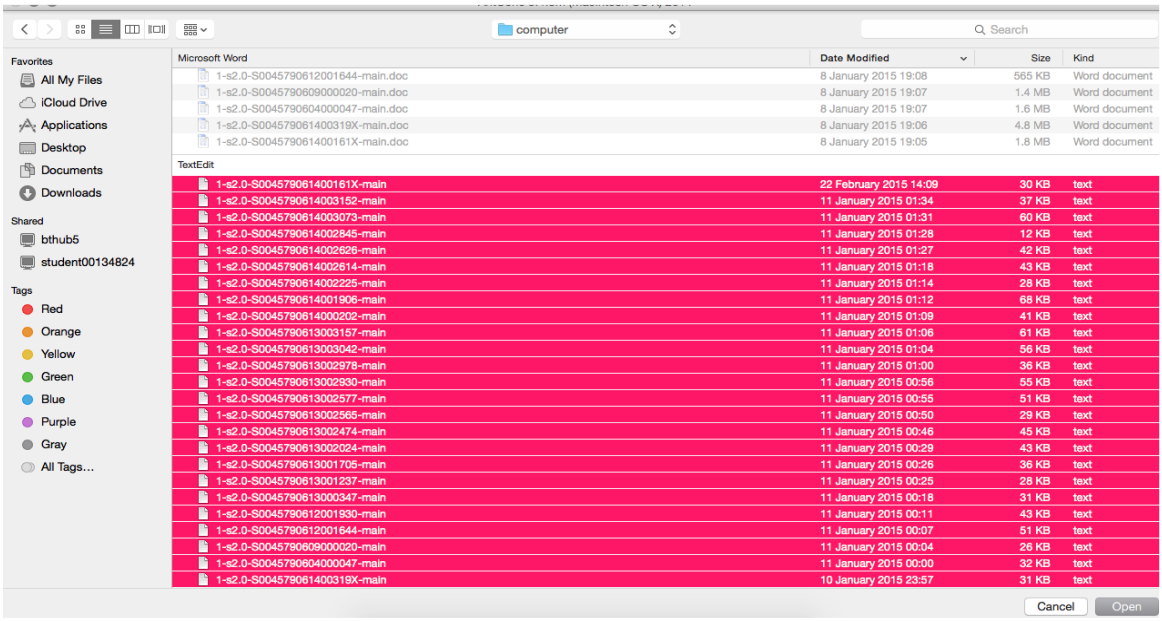


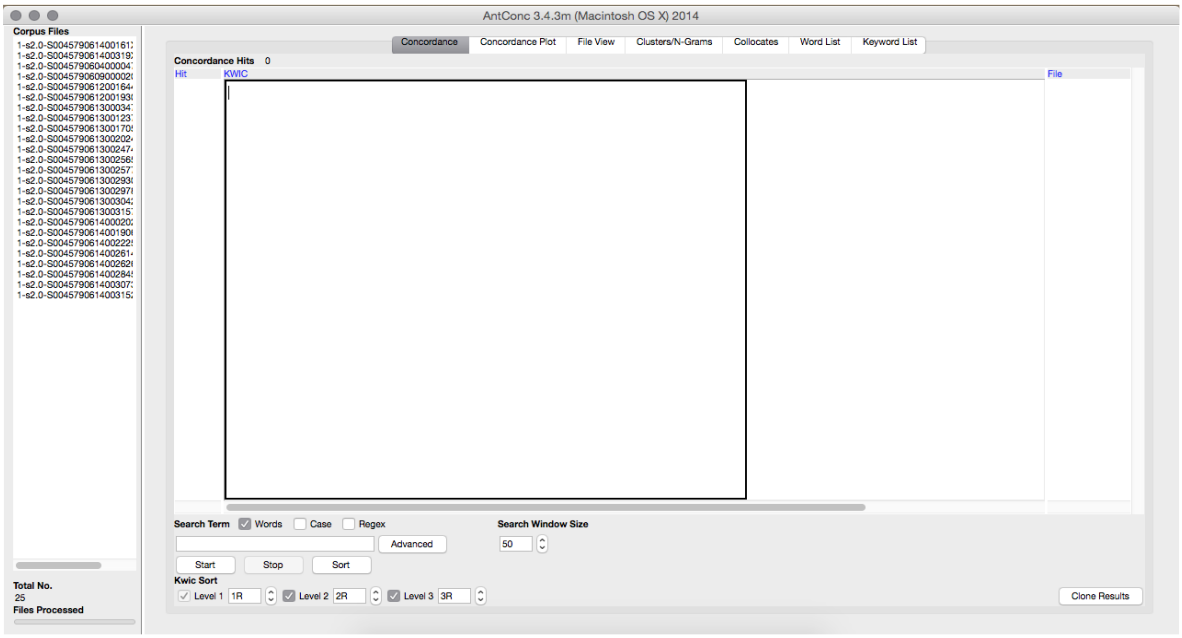
You can see two areas here corpus files and the results area (concordances hits), and you can see different tabs: concordance tool, concordance plot tool, file view tool, clusters/n-grams, collocates, word list, and keyword list. We will focus only on the concordance tool in this session.

Uploading ERAs corpus data

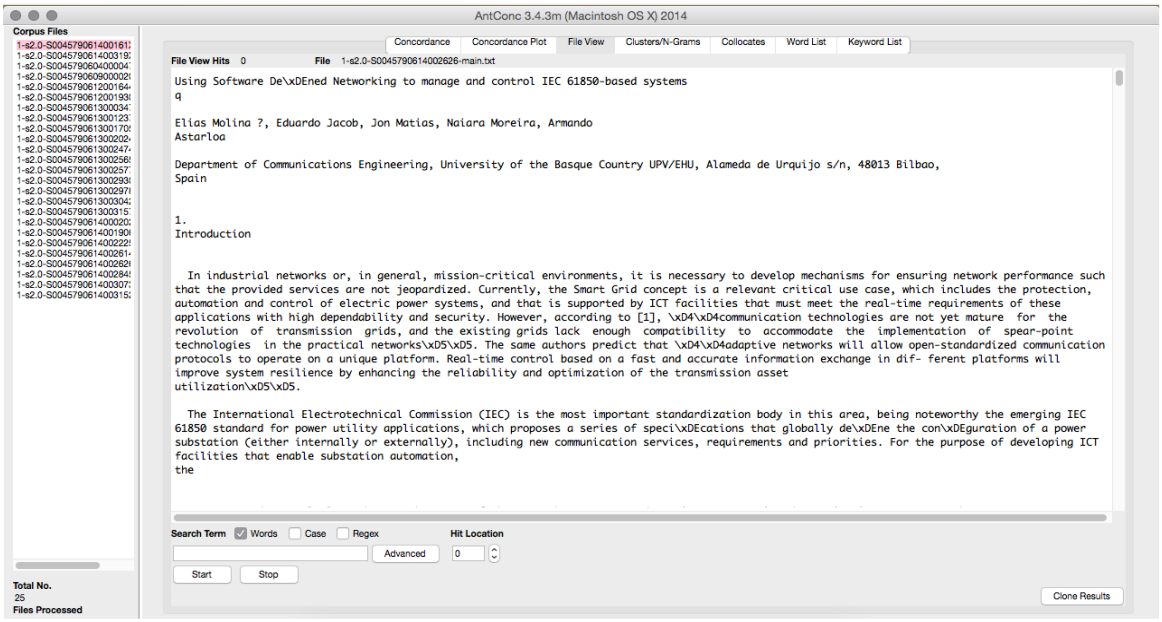
First we are going to click on file, and here you can open the file and navigate to your corpus.

You might have downloaded the ERAs corpus data in your computer. The ERAs corpus data are the files (in txt. format) I will shortly give it to you. After you have done downloading the ERAs file, keep and name the file in the folder where you can easily find. Open the file, drag and select the file you want or select Ctrl+A as following pages.





To make sure that the file is already uploaded, click on the first file to view whether you can see the file as the following page.



As you can see that it looks okay and you are now ready to do a search within the concordance tool.

Searching for a word and seeing its concordances

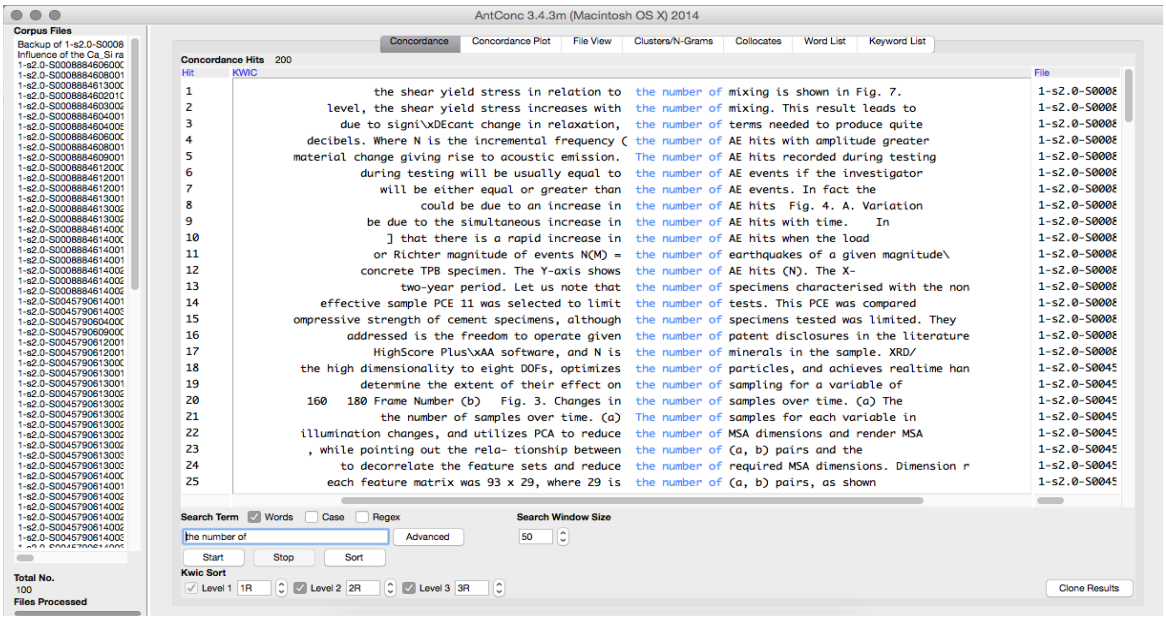
This tool shows search results in a 'KWIC' (KeyWord In Context) format. This allows you to see how words and phrases are commonly used in a corpus of texts.

Select one or more files for processing from using the 'Open File(s)...' or 'Open Dir...' options in the 'File' menu. The list of selected files is shown in the left frame of the main window.

Enter a search term on which to build concordance lines in the search box. First we are going to do the search on the fixed sequence of the word **the number of.** Type the word into the box below the search term and click on start.

You shoud be presented with the following page telling you there are 163 instances (hits) of the number of in the AntConc.

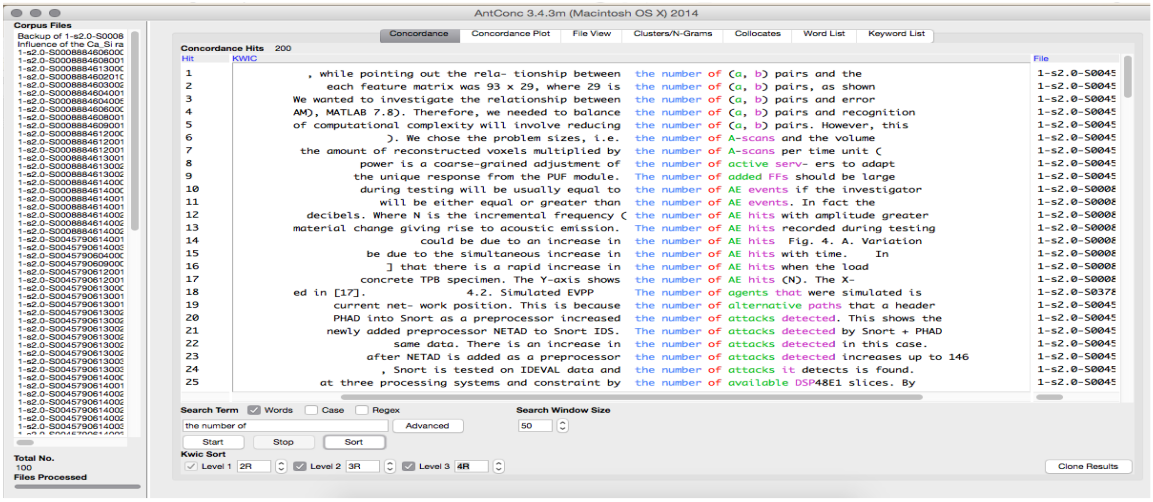
You can choose the number of text characters to be outputted on either side of the search term, using the increase and decrease buttons on the right of the button bar under the "Search Window Size" title (default value is 50 characters).



In this ordering that occurs here, it is difficult to see the pattern of **the number of**. If you want to see the pattern, it is good to sort the result. For example, by the word to the left and by the word to the right.

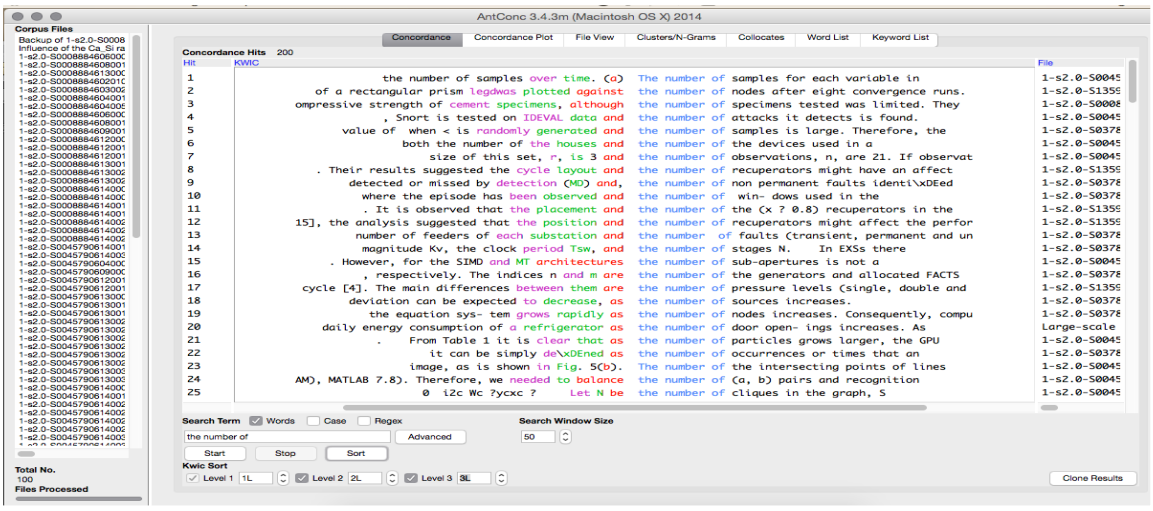
Use the Kwic Sort options to rearrange the concordance lines at three different levels. 0 is the search word, 1L, 2L... are words to the left of the target word, 1R, 2R... are words to the right of the target word.

Click on the **'Sort'** button to start the sorting process. At the moment, the concordance appear as the following page. You can scroll down the results to the 200 concordance lines. We can start to see the pattern now.

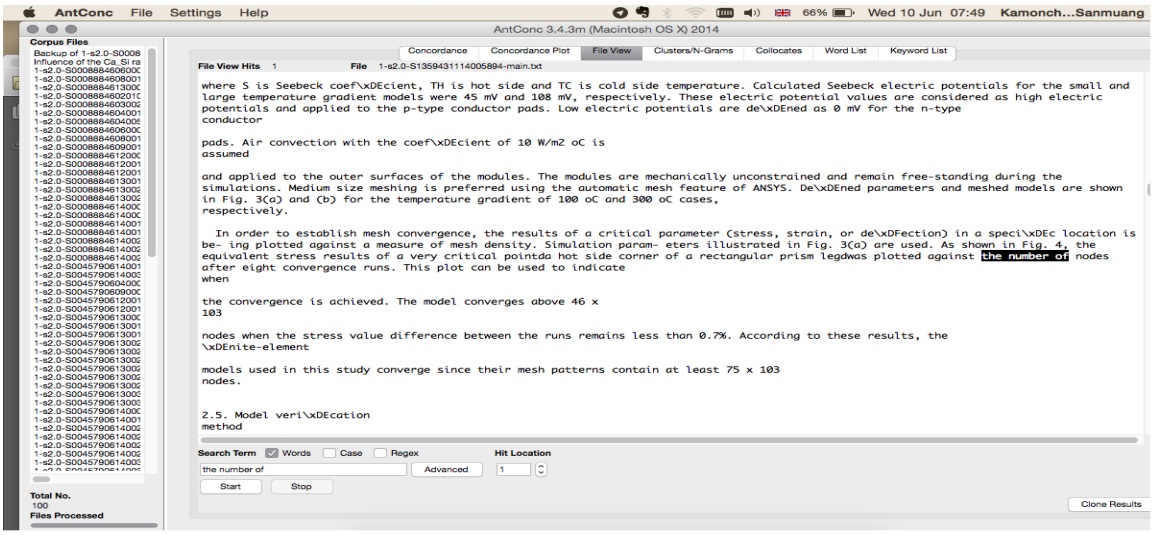


You can start to see the pattern of the target word now. From the first 25 examples, you can see that the number of is followed by the countable noun phrases in plural forms such as, *the number of A-scans* (lines 6 and 7)*, the number of AE hits* (from lines 12 to 17)*, the number of* agents (line 18).

You can also sort in different way, 1L, 2L, and 3L. You can see different kinds of pattern, you can see what kinds of word that come before **the number of** from the page below. You can see that the words come before the target word such as, *the connecting words* (against, as, and), *to infinitive form* (to balance), *be form* (be).



You can see any of these instances in more context. On the concordance screen hover the mouse over the word **the number of**, you can see a little finger here, and then click. You should be given a few paragraph of context in the original file (File View) (see below).



Now, you can start to search for another words. Please see the worksheet for the session 1 that I will shortly give it to you.

Activity 2 : Getting to know concordance lines and node words

Purposes: introduce concordance lines and node word (target word)

Instructions:

Do you know about concordance lines and node words (target words)? If not, scan the following lines of text taken from civil engineering research articles. Focus on the word 'important', and try to answer the questions that follow. You can discuss with the person near you. The answer will be explained and discussed together as a whole class.

Example

influence of the VEA content **is more important for** a greater time of rest.

of concrete is one of its **most important properties** and it is essential that the

and coarse aggregates. This **is an important observation** and will be analysed and

to Fig. 1b) and Table 3).The following **important observations** are obvious from the graphs

temperature is noted to be **the most important** one as it significantly alters the

The **following important conclusions** can be drawn from the

all products of hydration of cement; **the important resulting compounds** are calcium sulfate

or by applying a vacuum. While **extremely important, degassing conditions** are not always rep

w/cm ratio used and **it is important at** 28 days, the typical age for

caused by limestone filler. Thus, **it is important to** determine the proportion of limestone

of differential equations. Since they **are quite important**, the Dirichlet series representations are the soundness

of MgO concrete, **it is important to** employ accelerated experimental technique

expansion and cracking as one of **several important factors** for the deterioration.

(Type 30) compared to Type 10. **Another important parameter** in controlling the carbonation

grains and is usually characterised by **two important points** in hydration process, namely

is often considered as **the most important** period in the production of concrete structure

the cooling period which plays **an important role** in the shrinkage cracking at later

This type of work **is important** because it moves away from characterising

preferential and selective adsorption **is not only important for** mineral powder blends used in high As an important advantage over classical and widely used

the whole hydration period and presents **an important improvement** over the original procedure

The most important colouring, the one used to colour calcite

(Engineering Corpus)

Questions:

1. What is the target word or the search word that is under examination?
2. What is the verb that is always used before ‘important’? What are the other types of words used before ‘important’?
3. What preposition is most frequently used after ‘important’? What are the other types of words used after ‘important’?
4. What are the patterns of ‘important’ (e.g.,‘comparative form (more, the most) + important’)
5. What is the use of ‘important’ in the engineering research articles such as, describing experimental process ?
6. What is the potential meaning of ‘important’ in the text, positive or negative?

Now, you should know what concordance lines and node words are. If you are not sure, you can check with the possible answer on the next page

Possible answer:
The **concordance lines** above are lines of text taken from a corpus, i.e.,a collection of engineering research articles, which are organised and stored on a computer. The concordance lines may come from the beginning, the middle or the end of one of the texts. They may be made up of one sentence, part of a sentence or part of two sentences. You can see that each concordance line has a **node word** or the **target word**, the word that is being studied. Here we are studying the word 'important' that is used in the engineering research articles, or we study the words which are used before or after it.

Session 1:

Part 2: Finding lexical bundles

Activity A: where are the lexical bundles in the sentences?

Instructions:

1. From the nine sentences below, underline *the use of, the effect of, the number of*.

2. From the provided sentences, notice how each lexical bundle (*the use of, the effect of, the number of*) is placed and used in the sentences, such as in the beginning or the middle of the sentence, as a subject, or as an object.

If you are not sure what the subject or the object is, you can go to *self-study section* below in order to remind

1. The use of a local product depends on its specific properties and the requirements for a particular job.

2. If their claim is valid, then adequate protection of concrete should be ensured by the use of a low w/c alone or by the use of Type V cement alone.

3. The integration has been possible thanks to the use of a customized communication protocol, whose details of implementation are discussed. Finally Section 5 reports the conclusions.

4. It is instructive to see the effect of additive on creep and relaxation behavior.

5. The effect of Ca2+ intrusion reflects itself in the different gradings of SAP D.

6. A subroutine was written to consider the effect of conduction, convection and radiation during the laser deposition process

7. We analyse the influence of (a, b) pairs on MSA features as well as a range of suitable values, while pointing out the relationship between the number of (a, b) pairs and the recognition rate.

8. The number of added FFs should be large enough so that there is a high probability that PUF response sets the initial power-up state to one of the added states.

yourself of the subject and the object used in the sentences.

Self-study section: Subject and Object

Subject

The subject is a noun or a pronoun that comes before verb in an ordinary affirmative sentence. An affirmative sentence is often called active sentence or who or what does the action that the verb refers to (Adapted from Swan, 1996).

For example, *the engineer* is doing the experiment. This sentence is called active sentence, and the engineer is the subject and he does the action (experiment).

Object

The object can be also a noun or pronoun that normally comes after the verb, in an active clause. The direct object refers to a person or thing affected by the action of the verb, for example Take *the dog* for a walk (The dog is a direct object of this sentence). The indirect object usually refers to a person who receives the direct object e.g.,Ann gave *me a watch* (me is an indirect object and the direct object is a watch). (Adapted from Swan, 1996)

Activity 1: On your own, look at the words in **bold** whether they are the subject (S) or the object (O). Compare the answer with your friends.

1. **Radial basis function (RBFs) neural networks as a kind of powerful kernel methods** have been applied to **many areas** with success.
2. **The theoretical analysis of RBF structures and algorithms** includes **the orthogonal least square algorithm, the approximation capability analysis** [17,18,20]**, the design of RBF structure using fuzzy clustering method, the optimization of RBF structure using kernel orthonormalization method or combined supervised and unsupervised learning method, and the use of Fisher separability ratio for the selection of RBF centres**.
3. **The RBF** is selected because of **its compact support** [17,18,20].
4. Regarding Wiener, **it** has been implemented with **different structures for noise suppression applications** as MEG signals, acoustic signals, etc. [21].
5. **The effect of percentage of VPA on the permeability of VPC at the age of 28 days** is shown in **Fig. 2**.

Activity 2: Can you underline the subject and the object of the following sentences.
1. The permeability is increased from approximately 3.6 x 10 - 10 cm/s to around 13 x 10 - 10 cm/s when VPA content is increased from 0% to 100% by volume.
2. Compared to normal (0% VPA) concrete, the permeability of VPC with 100% VPA is about 3.5 times greater.
3. The changes in the number of equilibria now take place at higher values of the load.
4. The poorly damped oscillatory behaviour has been avoided producing a smoother operation.
5. In Eq. (9) the amount of lubricant delivered by the lubricator is calculated by counting the number of lubricant drops delivered in a given time period (usually 120 s).

What have you noticed from the above activities? We can say that the subject can be a single noun and can include all modifiers that go with it. And the object can be noticed by _____.

Activity B: Knowing the patterns of the lexical bundles

Instructions:

1. The following patterns tell us about how the lexical bundles (*the effect of, the number of*) are used in the research articles. Can you group the right concordances with the appropriate pattern?
2. When finished, you can share the answer with your neigbour and then you will be given the possible answer together as a whole class.
3. Can you guess which section of the research article each sentence comes from?

1. However, because the quantity of chloride ions absorbed by C–S–H gel is far lower than that bound by FS [42], *the effect of* decomposition of C–S–H gel on the stability of bound chlorides is insignificant when compared with the decomposition of FS under MgSO4 attack.

2. However, the values of radial stress at the inner and outer boundaries of the cylinder are not zero because of the effect of prescribed pressure loadings.

3. When we add a large number of new states |S0|, |S| + |S0| states can be implemented by a linear growth *in the number of* FFs that is log (|S| + |S0|).

4. Actually, when the dimensions of the generator are reduced, the induced voltage is decrease inherently as a result of decrease in the total flux even if *the number of* conductor remains the same.

5. Müllauer et al. investigated the effect of external sulfate attack on the leaching behavior of heavy metals in concrete.

6. Fig. 6 shows *the effect of* pile spacing on the thermally-induced mechanical behavior of energy pile (3*3 arrangement) in a group in sand.

7. This study has enabled investigating *the effect of* the water distribution system on the thermal performance of a forced draft counter-flow cooling tower (FDCT) filled with six different types of drift eliminators.

8. The only difference is that they were immersed in a similar covered container of 5% MgSO4 solution for 28, 56 and 90 days at the standard curing condition (20 ± 2 °C and 95% RH) for the examination of *the effect of* MgSO4 attack on the stability of bound chlorides.

9. Hajidavalloo et al. [18], in their work of cross flow cooling towers in variable wet-bulb temperature, include a brief discussion about *the effect of* the drift eliminator on tower performance, taking only a reduction of airflow rate into account.

10. The parameter ˇ is attributed in such a way that the total number of generated clones is twice *the number of* B cells in the population, i.e., nc = 200.
11. The effect of frequency and power was characterised on the depth and efficiency of concrete removal.

12. A reduction in *the number of* buffers improves the power efficiency; whereas, keeping a few necessary buffers intact, improves the performance.

13. In order to evaluate *the effect of* thermoelectric leg geometries on the power generation performance of thermoelectric modules, finite-element thermoelectric analyses were performed on rectangular-, trapezoidal-, octagonal-prism, and cylindrical models for the temperature gradient of 100 ºC.

14. *The number of* added FFs should be large enough so that there is a high probability that PUF response sets the initial power-up state to one of the added states.

15. *The number of* nodes can be reduced to improve efficiency because the aim of the simulation is to explore the relationship between the TF variations and the severity of the inter-turn fault rather than build a detailed numerical model for a specified winding.

Try doing the following exercise, assigning each of the sentences above to one of these patterns. **BUT**, if you are not sure what they are, there is an explanatory section about *clause* and *phrase* in the *self-study section* provided for you.

Pattern 1: beginning of a clause or phrase + *the effect of* or *the number of* as a dependent complement clause or prepositional

1.
2.
3.
4.

Pattern 2: beginning of a clause with *the effect of* or *the number of* embedded + a prepositional phrase

1.
2.
3.
4.
5.
6.

Pattern 3: begins with a *the effect of* clause + a prepositional phrase

1.

Pattern 4: begins with a phrase embedded with a *the effect of* or *the number of* clause plus a dependent complement clause or a prepositional phrase

1.
2.
3.
4.

Self-study section: Reviewing phrase and clause

A *phrase* consists of two or more words that function together as a group, such as a large long machine, in this study, the effect of percentage of VPA on the permeability of VPC at the age of 28 days.

In contrast, a *clause* contains a subject and verb, and it can convey a complete idea. The clause is joined to the rest of the sentence by a conjunction e.g.,*Mary said* that *she was tired*. There are two types of clause: an independent clause and dependent clause. An independent clause can express a complete thought (and can be a standalone sentence). A dependent clause is usually a supporting part of a sentence, and it cannot stand by itself as a meaningful proposition (idea).

The clause is sometimes contained participles or infinitives (with no subject or conjunction), such as *Not knowing what to do*, I telephoned Robin; I persuaded her *to try a new method.* You can see that the first sentence, '*Not knowing what to do*, I telephoned Robin.' contains present participle ending with '-ing' form (Not knowing what to do). The infinitive 'to + base verb' in the second one is 'to try a new method'. (Adapted from Swan, 1996)

Activity 1: On your own, identify independent and dependent clauses in the following sentences. Check the answer together in class.

1. Before beginning each rheological measurement, the mortar was agitated using a spoon in order to reduce the effect of possible sedimentation.
2. Thirdly, the use of the term "Na2O-equivalent" implies that the effect of potassium and sodium is equivalent.
3. However, there are numerous studies indicating that the influence of potassium and sodium on silica and quartz dissolution is not identical [e.g.,[18,19]].
4. Despite the work on the interaction of sulfate and chloride in concrete mentioned above, there is very little work on the effect of sulfate attack on the stability of bound chlorides in concrete.
5. In order to better capture the effect of the segment on the efficiency, the TE calculation is carried out at a constant load resistance condition.

Activity C: Knowing about engineering actions by the use of lexical bundles

Instructions:

1. Notice the use of the lexical bundles (*the use of, the effect of, the number of*) in the following concordance lines.
2. Underline the words that follow the lexical bundles i.e., *the effect of, the use of, and the number of*. Identify what types of word usually follow them such as, nouns, verbs, collocations, adverbs. Then, complete Table 1 below.
3. Do you think the words that co-occur with the lexical bundles mentioned play an important role in the sentence, and indicate specific actions in engineering? How?

1. The risk may then be compared with a standard defined by the utility or by a regulatory authority, in order to check whether or not it is necessary to increase *the number of* committed units in that situation.
2. *The effect of* percentage of VPA on the permeability of VPC at the age of 28 days is shown in Fig. 2.
3. However, *the use of* a high end computer and the wired installation per home increases the expense of the system.

4. *The use of* a multi-physic software plays a fundamental role for the design of power systems, due to that great part of the challenges are not limited to electrical and communication network but also involve other domain and energy networks (e.g.,heating and thermal) [7]

5. Ref. [12] involves *the use of* a networking cloud, Pachube, which is a real time data infrastructure that allows the management of data points from individuals, organizations and companies through the Internet.

6. The voltage divider model was used to study *the effect of* DG units on voltage dip.

7. Although there are detailed differences in the temperature distributions because of *the effect of* different convective cooling and the different disc geometry, overall the results demonstrated that the scaling methodology can be used with confidence for the design and development of automotive disc brake systems.

8. A reduction in the number of buffers improves the power efficiency; whereas, keeping a few necessary buffers intact, improves the performance.

9. This is because *the number of* alternative paths that a header can take to progress changes as it advances towards the destination.

10. If *the number of* faulty links surpasses the available spare links, then the split transmission and packet re-organization are performed.

Table 1

Item	Lexical bundles	Words that co-occur	Types of words	Indicate what
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				

Appendix (15). Revised Materials

Part 1: Presentation

Getting to Know the ERAs Corpus and Concordance Lines

- **What is a corpus?**

  A corpus (plural: corpora) is a collection of electronic texts compiled for specific purposes and accessed through text analysis software (Hunston, 2002; Charles, 2009).

  - Example: The *Engineering Research Articles Corpus (ERAs)* is designed to study lexical bundles in engineering research articles.

- **What is a concordance?**

  A concordance is a list of lines showing the context in which a specific word (called the "node" word) appears. It provides insights into word usage patterns in the corpus.

Activity 1: Exploring Concordance Lines

Focus: Understanding concordance lines.

Instructions: Examine concordance lines (Table 1) taken from the British Academic Written English Corpus (BAWE).

N concordance			
1	what was 'really going on' at the halfway	house	- what was 'really going on' was residents
2	'one-stop shop' for advice on benefits,	housing	, education, lifestyle, etc. They also provide
3	nationality, education, economics, distribution,	housing	, labour protection and social security
4	classic' welfare state: health, education,	housing	, and the national insurance and national
5	responsibilities, which include : education,	housing	, planning applications, strategic planning
6	level 24. This includes improving education,	housing	, material circumstances, targeting young
7	would be used for investments in education,	housing	, health care, or job training. Contrary
8	inequalities in many areas, including education,	housing	, and employment. Ethnic minorities are
9	that I expect to look at how the feeds,	housing	, fertiliser, etc. can be improved to meet
10	and environmental factors such as income,	housing	, education, safety, access to fresh food

- Answer the following questions as a group:
  1. What is the target word under examination?
  2. What types of words are used before and after the target word ("house")?
  3. What patterns of usage can you observe?
  4. What is the general meaning of the word "house" in these concordances?

Possible Answers (to discuss as a class):

1. The target word is "house," also called the "node" word.
2. Words before "house" are often nouns, while words after include adjectives or verbs.
3. Patterns of usage indicate "house" as an object in formal contexts.
4. The meaning of "house" varies; it can refer to a residence or a government system depending on the context.

Part 2: Practice

Comparing Concordance Lines Across Corpora

Activity 2: Analyse concordance lines from the *British Academic Spoken English Corpus (BASE)* (Table 2).

N concordance			
1	nineteenth century a novel called The Counting	House	[[voiced pause]] and by the way i use the
2	[[voiced pause]] in the [[voiced pause]] great	house	and then a younger servant a younger girl
3	it was within the compound of the great	house	surrounded by a high wall of Suffolk brick
4	Gladstone graveyard Anglican church plantation	house	my grandpa work so hard that he learn Latin
5	wood from the fence surrounding the great	house	or tools from the warehouse with the money
6	chapter from [[voiced pause]] this Counting	House	book right by the way i should say as a
7	[[voiced pause]] displaying them at in his	house	in London and allowing paying visitors
8	series [[voiced pause]] of discussions in the	House	of Commons through eighteen-fifteen and
9	estates with their rich and well equipped	houses	in the country and which was the worst
10	things like the woman being the angel in the	house	well most women in the nineteenth century

Instructions:

- Compare the use of the word "house" in *BASE* with its use in *BAWE* (Activity 1).
- Write your observations (in Thai or English) and share with the class.

Discussion Points:

- Does the usage differ in spoken vs. written corpora?
- Is the context more formal or informal in *BASE*?

Possible Answer:

In *BASE*, the word "house" often refers to a residence and is used in a more informal spoken context compared to the formal written context in *BAWE*.

Part 3: Production

Introduction to AntConc

Focus: Using AntConc to Explore the ERAs Corpus

- **Step-by-step Guide to Using AntConc:**
    1. Open AntConc and load the ERAs corpus.
    2. Search for a target word (e.g., "current").
    3. Sort concordance lines to identify patterns (left and right positions).

Activity 3: Interpreting Concordance Lines from the ERAs Corpus

Instructions:

- Use AntConc to search for the lexical bundle "current."
- Analyse concordance lines and identify:
    1. Patterns of usage.
    2. The meanings associated with "current" in the engineering context.
    3. The forms and functions of the bundle.

Reflection Task:

- Write your observations on the use of "current" in engineering research articles. Share how it differs from general academic contexts.