# STRUCTURAL AND FUNCTIONAL ANALYSES OF FOUR-WORD LEXICAL BUNDLES FROM ARTICLES IN THE FORBES WEBSITE


**BY**

**MISS LALITPATR RUJIRAWAN**


**AN INDEPENDENT STUDY PAPER SUBMITTED IN PARTIAL FULFILLMENT OF**
**THE REQUIREMENTS FOR THE DEGREE OF**
**MASTER OF ARTS IN**
**ENGLISH LANGUAGE TEACHING**
**LANGUAGE INSTITUTE**
**THAMMASAT UNIVERSITY**
**ACADEMIC YEAR 2020**


**COPYRIGHT OF THAMMASAT UNIVERSITY**



Ref. code: 25636221042135YPA

# STRUCTURAL AND FUNCTIONAL ANALYSES OF FOUR-WORD LEXICAL BUNDLES FROM ARTICLES IN THE FORBES WEBSITE

**BY**

**MISS LALITPATR RUJIRAWAN**

**AN INDEPENDENT STUDY PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS IN ENGLISH LANGUAGE TEACHING**

**LANGUAGE INSTITUTE**

**THAMMASAT UNIVERSITY**

**ACADEMIC YEAR 2020**

**COPYRIGHT OF THAMMASAT UNIVERSITY**

Ref. code: 25636221042135YPA

THAMMASAT UNIVERSITY
LANGUAGE INSTITUTE

INDEPENDENT STUDY PAPER

BY

MISS LALITPATR RUJIRAWAN

ENTITLED

STRUCTURAL AND FUNCTIONAL ANALYSES OF FOUR-WORD LEXICAL BUNDLES FROM ARTICLES IN FOREBS WEBSITE

was approved as partial fulfillment of the requirements for
the degree of Master of Arts in English Language Teaching

on May 30, 2021

Chair _____
(Associate Professor Supong Tangkiengsirisin, Ph.D.)

Member and Advisor _____
(Associate Professor Supakorn Phoocharoensil, Ph.D.)

Director _____
(Associate Professor Supong Tangkiengsirisin, Ph.D.)

Ref. code: 25636221042135YPA

| | |
|---|---|
| Independent Study Paper Title | STRUCTURAL AND FUNCTIONAL ANALYSES OF FOUR-WORD LEXICAL BUNDLES FROM ARTICLES IN THE FOREBS WEBSITE |
| Author | Miss Lalitpatr Rujirawan |
| Degree | Master of Arts |
| Major Field/Faculty/University | English Language Teaching<br>Language Institute<br>Thammasat University |
| Independent Study Paper Advisor | Associate Professor Supakorn Phoocharoensil, Ph.D. |
| Academic Year | 2020 |

# ABSTRACT

Lexical bundles are multiword sequences that commonly co-occur in a register (Bier & Conrad, 1999). Therefore, the lexical bundles used in various registers should be explored and investigated. The present study aimed to investigate and analyze the structures and discourse functions of lexical bundles employed in online articles regarding AI and 5G in Forbes.com, which belong to the journalism register. For these purposes, 362 articles under the AI and 5G subsections in Forbes.com were collected to build a corpus. The list of four-word lexical bundles was generated with the use of AntConc software. The structures of the bundles were analyzed and classified according to Biber et al.'s (2004) taxonomy, while the discourse functions were analyzed and classified employing Biber et al.'s (2004) and Hyland's (2008b) taxonomies. The results revealed that the most prevalent structure used in the journalism register was lexical bundles incorporating noun phrase and prepositional phrase fragments. Regarding the discourse functions, most bundles served as research-oriented bundles or referential expressions.

**Keywords**: Lexical Bundle Analysis, Corpus, Structural Analysis, Functional Analysis, Forbes.com

Ref. code: 25636221042135YPA

# ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my advisor, Associate Professor Supakorn Phoocharoensil, Ph.D., for all the support and guidance throughout my independent study process. He not only provided guidance but also inspired me to conduct this corpus-based study. I am also grateful to Associate Professor Supong Tangkiengsirisin, Ph.D. for his constructive comments and suggestions during my independent study defense.

I would also like to thank Mr. Theerat Chaweewan for taking the time to serve as the intercoder for this study.

Finally, I would like to express my deep gratitude to my parents, friends, and all my ELT classmates who have always supported and encouraged me throughout the MA study.

Miss Lalitpatr Rujirawan

Ref. code: 25636221042135YPA

# TABLE OF CONTENTS


| | Page |
|---|---|
| ABSTRACT | (1) |
| ACKNOWLEDGEMENTS | (2) |
| LIST OF TABLES | (6) |
| LIST OF FIGURES | (7) |
| LIST OF ABBREVIATION | (8) |
| CHAPTER 1 INTRODUCTION | 1 |
| 1.1 Background of the Study | 1 |
| 1.2 Research objectives | 3 |
| 1.3 Research questions | 3 |
| 1.4 Definition of terms | 3 |
| 1.5 Significance of the Study | 4 |
| 1.6 Scope of the Study | 4 |
| 1.7 Organization of the Study | 5 |
| CHAPTER 2 REVIEW OF LITERATURE | 6 |
| 2.1 Formulaic language | 6 |
| 2.2 Lexical bundles | 7 |
| 2.2.1 Structural types of lexical bundles | 7 |
| 2.2.2 Discourse functions of lexical bundles | 9 |
| 2.3 Previous studies | 13 |


Ref. code: 25636221042135YPA

CHAPTER 3 RESEARCH METHODOLOGY 17

3.1 Data collection 17
3.1.1 Forbes.com 17
3.2 Data analysis 18
3.2.1 The intercoder reliability 22
3.2.1.1 Intercoder selection 22
3.2.1.2 Intercoder training 22
3.2.1.3 Analysis of intercoder reliability 23
(1) Intercoder analysis of structural and functional types 23

CHAPTER 4 RESULTS AND DISCUSSION 24

4.1 Lexical bundles in the A5Fs corpus 24
4.2 Structural types of lexical bundles 28
4.3 Discourse functions of lexical bundles 33
4.3.1 Biber et al.'s (2004) discourse functions analysis 39
4.3.1.1 Stance bundles 39
4.3.1.2 Discourse organizers 41
4.3.1.3 Referential expressions 42
4.3.1.4 Special conversational functions 44
4.3.2 Hyland's (2008b) discourse functions analysis 44
4.3.2.1 Research-oriented 44
4.3.2.2 Text-oriented 45
4.3.2.3 Participant-oriented 46
4.4 Relationship between structural and functional categories of lexical bundles in the A5Fs corpus 47
4.5 Discussion 48

Ref. code: 25636221042135YPA

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS 50

5.1 Conclusion 50
5.2 Pedagogical implications 50
5.3 Limitations and Recommendations 52

REFERENCES 54

APPENDICES

APPENDIX A 62
APPENDIX B 64

BIOGRAPHY 66

Ref. code: 25636221042135YPA

# LIST OF TABLES


| Tables | Page |
|---|---|
| 1 Structural types of spoken and written lexical bundles in academic discourse (Biber et al., 2004) | 8 |
| 2 Discourse functions of spoken and written lexical bundles in academic discourse (Biber et al., 2004) | 10 |
| 3 Functions of bundles in research writing (Hyland, 2008b) | 12 |
| 4 Top-133 most Lexical Bundles in the A5Fs Corpus | 24 |
| 5 Structural distribution of lexical bundles in the A5Fs corpus | 29 |
| 6 Structural analysis of lexical bundles in the A5Fs corpus | 31 |
| 7 Discourse functions distribution of lexical bundles in the A5Fs corpus | 34 |
| 8 Discourse functions analysis of lexical bundles in the A5Fs corpus according to Biber et al.'s (2004) taxonomy | 36 |
| 9 Discourse functions analysis of lexical bundles in the A5Fs corpus according to Hyland's (2004) taxonomy | 38 |


Ref. code: 25636221042135YPA

# LIST OF FIGURES


| Figures | Page |
|---|---|
| 1 Total word tokens derived with Word List feature of AntConc | 19 |
| 2 Data analysis undertaken with Clusters/N-Grams feature of AntConc | 20 |
| 3 4-gram tokens extracted with Clusters/N-Grams feature of AntConc | 20 |
| 4 5-gram tokens extracted with Clusters/N-Grams feature of AntConc | 21 |
| 5 Relationship between the structural and functional categories in the A5Fs corpus | 48 |


Ref. code: 25636221042135YPA

# LIST OF ABBREVIATIONS

| Symbol/Abbreviation | Term |
|---|---|
| A5Fs | The articles under AI and 5G subsections in Forbes.com |

Ref. code: 25636221042135YPA

# CHAPTER 1
# INTRODUCTION

## 1.1 Background of the Study

At present, English is the medium used to communicate worldwide, with the number of English users across the world predicted to reach almost two billion people (Crystal, 2008a). English writing skill is necessary in the global stage for three reasons: the increasing number of international linguists with writing as their specialization, the proliferation of English journals or articles, and the growing number of international students who are pursuing studies in English-speaking countries (Santos et al., 2000). Meanwhile, the rising number of English journals or articles is evidence that English has become the dominant language of intellectual communities (Short, Boniche, Kim & Li, 2001). In 2019, Google Analytics showed that the Bangkok Post's website was accessed by around 16.7 million people and reached 84.7 million views, which increased by 11.9% from 2018 (Bangkok Post, https://www.bangkokpost.co.th/business/newspaper/bangkokpost). The average page views daily was 230,000 and accessed by 91,206 devices. These numbers show that a growing number of people read news online. The online platform of news is also easy to access through smartphones or computers, anytime and anywhere.

Lexical bundles are multiword sequences which frequently occur in natural discourse, regardless of their idiomaticity and structural status, and are distributed across different texts (Biber et al., 1999; Biber, 2006a; Biber, 2010; Hyland, 2012). Lexical bundles are stored and can be retrieved from the mental lexicon in language production (Nekrasova, 2009). The frequent occurrence of these multiword sequences aid both language production and reception (Rafiee et al., 2011). Erman (2007), Pawley & Syder (1983), Raupach (1984) and Wood (2006) agreed that it minimizes the process of encoding and decoding messages in spoken and written discourse. It has been stated that the appropriate usage of lexical bundles in communication can help language learners reach near native-like proficiency and fluency, and enhance the understanding of meaning and context in particular contexts or discourse (Byrd & Coxhead, 2010;

Ref. code: 25636221042135YPA

Hong & Hua, 2018). Jalali and Moini (2014) agreed that these multiword sequences are built and used based on specific purpose or some technical words. Reid (1990) also contends that the nature of typical writing tasks or topics affects writers' use of cohesive devices, word choice and performance. The results from previous studies have shown that lexical bundles can help to shape the meaning of a text and the distinctiveness in different registers, disciplines and text types (Hyland, 2008b; Rafiee et al., 2011). These frequently occurring fixed expression can enable learners to gain communicative competence (Hyland, 2008b). Therefore, there are many studies of lexical bundles that incorporated the use of corpus analysis in the past few decades which aimed to identify how they function across different disciplines.

This study aims to investigate the function and structure of four-word lexical bundles used in the Forbes website (Forbes.com) in the AI, artificial intelligence, and 5G subsections. Despite the increasing number of studies on lexical bundles, few studies have investigated the functions of lexical bundles used in media such as Forbes. The reason for investigating the lexical bundles used in Forbes.com is that nowadays people tend to read articles or news online, as the online platform of media can be updated more frequently and costs less than the offline one. Technology also plays a vital part in everyday life and most people rely heavily on it (The Importance of Technology in Our Daily Life – How Has Technology Changed Our Lives?, https://www.scientificworldinfo.com/2019/11/importance-of-technology-in-our-daily-life.html). Moreover, the articles in Forbes.com are written by both native English writers and international contributors, unpaid writers and domain experts, and their articles are reviewed by Forbes' editors before they are accepted as contributors and their articles are posted online (Steimle, n.d.). Even though some articles were written by international contributors, they have been reviewed, accepted and sometimes got edited by editors at Forbes (Steimle, n.d.). Hence, the lexical bundles used in these articles will reflect natural and authentic use of the English written genre. Therefore, the results of this study will be valuable for both English language teachers who aim to conduct a class for journalism students, journalism students who aim to work in this field, and journalists who wish to develop their writing skills.

Ref. code: 25636221042135YPA

## 1.2 Research Objectives

1) To investigate the top-133 most frequent lexical bundles in AI and 5G subsections in Forbes.com (A5Fs).
2) To examine the structural types of the top-133 most frequent lexical bundles used in A5Fs.
3) To examine the discourse functions of the top-133 most frequent lexical bundles used in A5Fs.

## 1.3 Research Questions

1) What are the top-133 most frequent lexical bundles in AI and 5G subsections in Forbes.com (A5Fs)?
2) What are the structural types of the top-133 most frequent lexical bundles used in A5Fs?
3) What are the discourse functions of the top-133 most frequent lexical bundles used in A5Fs?

## 1.4 Definition of Terms

1.4.1) AI (artificial intelligence) is the science and engineering of making intelligent programs or machines which are capable of performing particular tasks (McCarthy, 1998). It is being used in various industries. Any machine with human mind traits such as learning and problem-solving is also AI (Frankenfield, 2021). Some examples of AI are self-driving cars, Apple's Siri, Google assistant or applications in the financial industry which are used to flag unusual activity. All these AIs have the ability to learn, analyze or solve problems.

1.4.2) 5G is the global wireless fifth generation of mobile network after 1G, 2G, 3G and 4G networks. It is designed to virtually connect everyone and everything including machines, objects and devices (What is 5G?, https://www.qualcomm.com/5g/what-is-5g)

Ref. code: 25636221042135YPA

1.4.3) Forbes is a global media and publishing company which was founded in the United States in 1917 (Kenton, 2021). Forbes provides daily news focusing on business, investment, technology, entrepreneurship, leadership and lifestyle (Kenton, 2021).

1.4.4) Lexical bundles, in this study, are used to describe fixed or semi-fixed four-word bundles that commonly occur together in Forbes.com.

## 1.5 Significance of the Study

For this study the researcher collected 254 articles from the Forbes website in the AI subsection from January 12 – April 3, 2021 and 108 articles in the 5G subsection from June 16, 2020 – March 31, 2021. The articles were used to explore and analyze the structural and discourse functions of four-word lexical bundles written by native speakers of English and the international contributors of Forbes.com. The results from this study will be beneficial for pedagogical use for English teachers, enabling them to introduce these four-word lexical bundles to learners in a journal writing class. English teachers can also implement them as a guideline for course development, teaching strategies, and authentic teaching material to enhance learners' journal writing skills in the journal writing classroom. Moreover, journalists can also benefit from the results of this study in terms of developing and mastering their writing skills.

## 1.6 Scope of the Study

The scope of this study is to investigate the articles from Forbes website under the AI and 5G subsection from June 2020 to April 2021. This study mainly explored the structural types and the functions of lexical bundles in the A5Fs corpus. However, this study did not include contrastive analysis of articles written by native English and international writers. This study employed only one specific framework to analyze the structures of lexical bundles, i.e. Biber et al.'s (2004) taxonomy. Meanwhile, Biber et al.'s (2004) and Hyland's (2008b) taxonomies were employed to analyze the functions of lexical bundles in this study.

## 1.7 Organization of the Study

Ref. code: 25636221042135YPA

The present study consists of five chapters:

1.7.1) Chapter one is the introduction of the study. This chapter provides an overview of the study which covers the background of the study, the research objectives, the research questions, the definition of terms, the significance of the study, the scope of the study, and the organization of the study.

1.7.2) Chapter two provides the relevant previous studies which cover the definition and explanations of formulaic language and lexical bundles. The framework for the functional and structural analysis of lexical bundles is reviewed and presented in this chapter. Moreover, the previous studies on lexical bundles comparing the use of bundles between L1 and L2 learners, advanced and student writers, and between different levels of study have been reviewed. This chapter also includes a review of the studies on the relationship between structural and functional categories of lexical bundles and the use of lexical bundles in different disciplines.

1.7.3) Chapter three proposes the research methodology including data collection, data analysis of the target corpus and the intercoder reliability. This chapter also described the concordance software, i.e. AntConc (Anthony, 2020), used to extract the lexical bundles.

1.7.4) Chapter four presents the results of the study: functional categories and structural types of the four-word lexical bundles in the target corpus. The lists of lexical bundles and the relationship between structural and functional categories of lexical bundles in the target corpus are also presented and discussed.

1.7.5) Chapter five contains the conclusion of the study, the implications and recommendations for further study. This chapter also details the limitations of the study.

Ref. code: 25636221042135YPA

# CHAPTER 2
# REVIEW OF LITERATURE

## 2.1 Formulaic Language

Formulaic language is a term describing combinations of words that fulfill specific functions and recalled mostly automatically as a chunk, like a single lexical item, by native speakers (Schmitt & Carter, 2004; Wood, 2002). This formulaic language can consist of one word to many words, which can be a repetition of fixed and semi-fixed multiword combinations (Byrd & Coxhead, 2010). It is segmented from input and stored as a whole in long-term memory, which is fundamental to produce fluent or near native-like language (Wood, 2002). It has been indicated that every genre incorporates specific expressions which identify that genre (Jalali & Moini, 2014). Therefore, the importance of multiword items in spoken and written discourse in different genres have been emphasized in several studies (Biber & Cortes, 2004; Cowie, 1992; Moon, 1998). Studies have shown that these multiword items, or formulaic language units, are significant both in language acquisition and production (Wood, 2002). The use of this combination of words demonstrates membership in a particular discourse community: “when we speak, we select particular turns of phrase that we perceive to be associated with certain values, styles and groups” (Wray, 2006).

It has been found that adult learners do not extensively use prefabricated formulaic language (Yorio, 1989). However, when they used the formulaic language in their writings, it appears that they do not use it to further their language development but rather as a production strategy (Yorio, 1989). Howarth (1998) mentioned one difficulty that non-native academic writers may encounter is the use of appropriate verbs with specific nouns, which may lead to miscommunication. The study of Durrant and Schmitt (2009) also showed that non-native speakers overuse high-frequency collocations, which indicates significant differences between native and non-native writers. It has also been shown that non-native writers have poorer intuitions about particular collocations (Siyanova & Schmitt, 2008).

Ref. code: 25636221042135YPA

## 2.2 Lexical Bundles

Lexical bundles, the multi-word sequences that occur most commonly in a particular register, are the most frequently recurring sequences of words. The term 'lexical bundles' was first used in the Longman Grammar of Spoken and Written English (Biber et al., 1999) in regard to comparing the most common recurrent sequences in conversations and academic prose. Conrad and Biber (2005) have classified lexical bundles in two major ways, namely the structural characteristics of bundles and their function in a discourse context.

### 2.2.1 Structural Types of Lexical Bundles

Even though lexical bundles are sometimes structurally incomplete, "they have strong grammatical correlates" (Biber et al., 2004). Biber et al. (2004) identified three major structural types of lexical bundles, shown in Table 1.

The first type incorporates verb phrase (VP) fragments and consists of seven sub-types. A verb phrase fragment in types 1a and 1b follows a subject pronoun, be it first-, second- or third-person pronoun (e.g. *that's one of the*). A verb phrase in type 1c follows a discourse marker (e.g. *I mean I don't*). A verb phrase in types 1d and 1e simply begins the bundle (e.g. *have a lot of, can be used to*), while a verb phrase in types 1f and 1g is a question fragment (e.g. *do you want to, what does that mean*).

The second type incorporates dependent clause fragments with simple verb fragments. Type 2a bundles consist of complementizer, a word used to introduce a complement clause (e.g. *to, if*), which follows a main clause (e.g. *you might want to, I don't know if*). Types 2b-2e are dependent clause fragments which begin with a complementizer or subordinator, a word used to introduce a subordinate clause (e.g. *if you want to, to come up with, that this is a, when we get to*).

Type 1 and type 2 bundles have clausal components. However, type 3 bundles are phrasal. Types 3a-3c consist of noun phrase components. Type 3a begins with a noun phrase, sometimes with a connector, followed by *of*-phrase fragment (e.g. *one of the thing, a little bit of*). Type 3b begins with a noun phrase followed by a postmodifier fragment other than of (e.g. *a little bit about, the way in which*). Type 3c consists of noun phrase expressions (e.g. *or something like that*). Type 3d incorporates

Ref. code: 25636221042135YPA

prepositional phrase expressions (e.g. *at the same time*), while type 3e consists of comparative expressions (e.g. *as far as the*).

**Table 1.** Structural types of spoken and written lexical bundles in academic discourse (Biber et al., 2004)

| **Structural types** | **Sub-types** | **Sample bundles** |
|---|---|---|
| **1.Lexical bundles that incorporate *verb phrase* fragments** | 1a. (connector +) 1st / 2nd person pronoun + VP fragment | *you don't have to* |
| | | *well I don't know* |
| | | *I'm not going to* |
| | 1b. (connector +) 3rd person pronoun + VP fragment | *it's going to be* |
| | | *that's one of the* |
| | | *and this is a* |
| | 1c. Discourse marker + VP fragment | *I mean you know* |
| | | *you know it was* |
| | | *I mean I don't* |
| | 1d. Verb phrase (with non-passive verb) | *is going to be* |
| | | *is one of the* |
| | | *have a lot of* |
| | 1e. Verb phrase with passive verb | *is based on the* |
| | | *can be used to* |
| | | *shown in figure N* |
| | 1f. *yes-no* question fragments | *are you going to* |
| | | *do you want to* |
| | | *does that make sense* |
| | 1g. WH-question fragments | *what do you think* |
| | | *how many of you* |
| | | *what does that mean* |

Ref. code: 25636221042135YPA

**Table 1.** (Continued)

| Structural types | Sub-types | Sample bundles |
|---|---|---|
| **2.Lexical bundles that incorporate *dependent clause* fragments** | 2a. 1st / 2nd person pronoun + dependent clause fragment | *I want you to* |
| | | *you might want to* |
| | | *I don't know if* |
| | 2b. WH-clause fragments | *what I want to* |
| | | *what's going to happen* |
| | | *when we get to* |
| | 2c. *If*-clause fragments | *if you want to* |
| | | *if we look at* |
| | | *if you have a* |
| | 2d. (verb/adjective +) *to*-clause fragment | *to be able to* |
| | | *to come up with* |
| | | *want to do is* |
| | 2e. That-clause fragments | *that there is a* |
| | | *that I want to* |
| | | *that this is a* |
| **3.Lexical bundles that incorporate *noun phrase and prepositional phrase* fragments** | 3a. (connector +) Noun phrase with *of*-phrase fragment | *one of the things* |
| | | *the end of the* |
| | | *a little bit of* |
| | 3b. Noun phrase with other post-modifier fragment | *a little bit about* |
| | | *those of you who* |
| | | *the way in which* |
| | 3c. Other noun phrase expressions | *a little bit more* |
| | | *or something like that* |
| | | *and stuff like that* |
| | 3d. Prepositional phrase expressions | *of the things that* |
| | | *at the end of* |
| | | *at the same time* |
| | 3e. Comparative expressions | *as far as the* |
| | | *greater than or equal* |
| | | *as well as the* |

### 2.2.2 Discourse Functions of Lexical Bundles

Lexical bundles are sometimes structurally incomplete and not idiomatic in meaning, but they serve important discourse functions in both spoken and written texts (Biber & Barbieri, 2007). The discourse functions of spoken and written lexical bundles

Ref. code: 25636221042135YPA

in academic discourse introduced by Biber et al. (2004) consist of four main categories: stance bundles, discourse organizers, referential expressions, and special conversation functions, as shown in Table 2.

The stance bundles are used to convey the speaker's attitude, judgement, perspective, certainty or uncertainty. Stance bundles are further divided into two main sub-categories: epistemic stance and attitudinal/modality stance. Epistemic stance provides the information of certainty, uncertainty or possibility. Attitudinal/Modality stance delivers speaker's attitudes towards actions or events, which is further divided into four sub-categories: desire, obligation/directive, intention/prediction, and ability. Discourse organizers are used to signal both before and after discourse, which consist of topic introduction and topic elaboration. Referential expressions are likely to provide support to physical objects, abstract ideas, places, time, text or multifunction. The special conversational functions are used to introduce politeness, enquiry, and reporting clause.

**Table 2.** Discourse functions of spoken and written lexical bundles in academic discourse (Biber et al., 2004)

| **Categories** | **Sub-categories** | **Examples** |
|---|---|---|
| **1. Stance bundles** | 1.1 Epistemic stance | *I don't know if* |
| | | *I don't know how* |
| | 1.2 Attitudinal/modality stance | |
| | Desire | *if you want to* |
| | | *do you want a* |
| | Obligation/directive | *I want you to* |
| | | *You need to know* |
| | Intention/prediction | *I'm going to* |
| | | *it's going to* |
| | Ability | *to be able to* |
| | | *can be used to* |
| **2. Discourse organizers** | 2.1 Topic introduction | *what do you think* |
| | | *going to talk about* |
| | 2.2 Topic elaboration/clarification | *has to do with* |
| | | *you know I mean* |

Ref. code: 25636221042135YPA

**Table 2.** (Continued)

| Categories | Sub-categories | Examples |
|---|---|---|
| **3. Referential expressions** | 3.1 Specification of attributes | |
| | Intangible framing attributes | *the nature of the* |
| | | *in case of the* |
| | Tangible framing attributes | *the size of the* |
| | | *in the form of* |
| | Quantity | *there's a lot of* |
| | | *the rest of the* |
| | 3.2 Identification/focus | *that's one of the* |
| | | *those of you who* |
| | 3.3 Imprecision | *or something like that* |
| | | *those of you who* |
| | 3.4 Time/place/text reference | |
| | Place reference | *The United States and* |
| | | *of the United States* |
| | Time reference | *at the same time* |
| | | *at the time of* |
| | Text deixis | *shown in figure N* |
| | | *as shown in the figure* |
| | Multifunctional reference | *in the middle of* |
| | | *the end of the* |
| **4. Special conversational functions** | 4.1 Politeness | *thank you very much* |
| | 4.2 Simple inquiry | *what are you doing* |
| | 4.3 Reporting | *I said to him/ her* |

According to Hyland (2008b), Biber's (Biber, 2006b; Biber et al., 2004) framework emerged from a broader corpus of written and spoken registers, which may lead to an unproductive comparison when grouping bundles under the same sub-categories. Therefore, Hyland (2008b) developed a framework to analyze the discourse functions of lexical bundles from Biber's (Biber, 2006b; Biber et al., 2004) taxonomy that focused more on research writing. Hyland (2008a) collected bundles which fall into three categories: research-oriented, text-oriented, and participant-oriented as shown in Table 3. These broad classifications aimed to minimize the overlaps between categories

Ref. code: 25636221042135YPA

(Hyland, 2008a). Sub-categories of each category were introduced to explicitly display the concerns of research writing (Hyland, 2008b).

Research-oriented bundles convey writers' real-world activities and experiences which serve an ideational function. This category indicates and explains location, procedure, quantification, description of research objects, and topic related to the research (Hyland, 2008a). Text-oriented bundles refer to the organization of a text and its meaning as a message or argument (Hyland, 2008a). These bundles are used to provide ways to engage with a literature, provide warrants, connect ideas, direct readers around the text, and specify limitations (Hyland, 2008b). Text-oriented bundles consist of transition signals, resultative signals, structuring signals, and framing signals. They are used to refer to the organization of the text and its meaning. Participant-oriented bundles 'focus on the writer or reader of the text' (Hyland, 2005). These bundles convey the writers' attitude and evaluation through stance features and directly address readers through engagement features (Hyland, 2008a). However, each register employs a particular set of lexical bundles, together with the communicative purposes of that register.

**Table 3.** Functions of bundles in research writing (Hyland, 2008b)

| **Categories** | **Sub-categories** | **Examples** |
|---|---|---|
| **1. Research-oriented** | Location | *at the beginning of* |
| | | *at the same time* |
| | | *in the present study* |
| | Procedure | *the use of the* |
| | | *the role of the* |
| | | *the purpose of the* |
| | Quantification | *the magnitude of the* |
| | | *a wide range of* |
| | | *one of the most* |
| | Description | *the structure of the* |
| | | *the size of the* |
| | | *the surface of the* |
| | Topic | *in the Hong Kong* |
| | | *the currency board system* |

Ref. code: 25636221042135YPA

**Table 3.** (Continued)

| Categories | Sub-categories | Examples |
|---|---|---|
| **2. Text-oriented** | Transition signals | *on the other hand* |
| | | *in addition to the* |
| | | *in contrast to the* |
| | Resultative signals | *as a result of* |
| | | *it was found that* |
| | | *these results suggest that* |
| | Structuring signals | *in the present study* |
| | | *in the next section* |
| | | *as shown in figure* |
| | Framing signals | *in the case of* |
| | | *with respect to the* |
| | | *on the basis of the* |
| | | *in the presence of* |
| | | *with the exception of* |
| **3. Participant-oriented** | Stance features | *are likely to be* |
| | | *may be due to* |
| | | *it is possible that* |
| | Engagement features | *it should be noted that* |
| | | *as can be seen* |

## 2.3 Previous Studies

Various studies have compared and analyzed the use of lexical bundles between English native speakers and non-native writers and learners. The results have shown that non-native writers "rely on more limited set of lexical phrases" and often use the same "safe" bundles in their writing (Juknevičien, 2009). Pang (2009) determined that non-native learners use different lexical bundles from those of native speakers of English. Non-native learners also "use four times as many lexical bundles as the native speakers do" (Pang, 2009). Hyland (2008b) also found that the non-native graduate students applied wider a number and range of four-word bundles than published writers did.

Chen and Baker (2010) examined both structural and functional categories of four-word lexical bundles from one corpus of published academic texts and two corpora

Ref. code: 25636221042135YPA

of student academic writing. The first corpus was a corpus of native academic writers, whereas the other two corpora were corpora of native and Chinese (non-native) university students. The results showed that the use of lexical bundles in the writing of student writers appears to be a sign of immature writing (Chen & Baker, 2010). The structural types which student writers tended to use in the writing were VP-based bundles, while native academic writers used more NP-based bundles (Chen & Baker, 2010). The results from other studies also showed that native speakers of English tend to use NP-based bundles more than non-native learners do (Pan et al., 2015; Güngör & Uysal, 2016). Moreover, non-native students did not use the VP with the passive verb form as frequently as native speakers of English did, and those VP with passive verb bundles were not used by both groups of native writers (Chen & Baker, 2010). The results also showed that native academic writers and student writers relied on different functional lexical bundles (Chen & Baker, 2010). Native academic writers relied heavily on referential bundles, while student writers relied more on discourse organizers (Chen & Baker, 2010). Furthermore, non-native student writers tend to use certain deictic expressions which may be identified as informal or over-generalizing expressions, and they even misuse certain bundles (Chen & Baker, 2010; Pan et al., 2015).

There is a strong relationship between the structural and functional categories of lexical bundles (Biber et al., 2004). It was found that "most stance bundles are composed of dependent clause, while most referential bundles are composed of noun phrase or prepositional phrase fragments" (Biber et al., 2004). However, the results from the study of Hyland (2008a) showed that research articles contain the most discourse organizers and stance bundles and the fewest referential bundles. The reason is that there is new knowledge presented as well as elicitation of peers' recognition in research articles (Hyland, 2008a). Cortes (2004) also states that students' writing in higher levels of study employed more organizational patterns of bundles because the writing tasks are more research-oriented than that of lower levels of study.

It is noteworthy that the proficiency level of learners also correlates with the number and variety of lexical bundles used in their writing (Qin, 2014). Non-native academic writers at the higher level of study also use more academic writing bundles (e.g., *NP with post-modifier fragments* or *prepositional phrases*) and discourse

Ref. code: 25636221042135YPA

organizers (e.g., *explanation*, *exemplification*, and *focus*) than non-native academic learners in a lower level of study (Qin, 2014). The inherently complex structural forms of *NP with post-modifier fragments*, *prepositional phrases* or *past participle* phrases, which is an important feature of academic writing, may influence how different levels of non-native graduate writers employ these structural types of bundles in their writing (Biber et al., 1999: Qin, 2014).

Lexical bundles in various registers and genres have been studied for decades. Most of the studies on written registers have focused on academic discourse (Cortes, 2004; Hyland, 2008b; Juknevičien, 2009; Pang, 2009; Chen & Baker, 2010; Llanta, 2014; Qin, 2014; Pan et al., 2015; Güngör & Uysal, 2016). However, the structural and functional types in the journalism register (Rafiee et al., 2011; Dastjerdi & Rafiee, 2013; Fadhilah, 2020) have not been explored as much as other registers. The findings have revealed that the structural types of lexical bundles which are used in the journalism registers are mostly aligned with phrase-based bundles in academic writing (Dastjerdi & Rafiee, 2013). Both native English and non-native journalists employed more phrasal lexical bundles than clausal lexical bundles (Rafiee et al., 2011). Regarding the functional categories of lexical bundles in this register, it has been found that the referential bundles are more frequent than the other categories (Dastjerdi & Rafiee, 2013; Rafiee & Keihaniyan, 2013; Fadhilah, 2020). The results are aligned with Chen and Baker (2010), Güngör and Uysal (2016), and Pan et al.'s (2016) studies showing that native English writers rely heavily on referential or research-oriented bundles.

A number of corpus-based studies have examined the use of lexical bundles in different disciplines, and the results showed that there are different and unique recurrent word combinations in each discipline or academic community (Qin, 2014). Furthermore, lexical bundles have been shown to be discipline-based (Cortes, 2004; Hyland, 2008b; Jablonkai, 2010). Llanta (2014) compared the use of lexical bundles in expert academic writing with three language variables—L1 English, L2 English and L1 Spanish—in twelve different disciplines. The results demonstrated that the use of lexical bundles is determined by register (Llanta, 2014). Therefore, it is suggested that corpus-based studies in different disciplines should be conducted to explore the use of

Ref. code: 25636221042135YPA

and the most frequent lexical bundles to benefit students in those specific majors (Qin, 2014).

Ref. code: 25636221042135YPA

# CHAPTER 3
# RESEARCH METHODOLOGY

## 3.1 Data Collection

In this study, articles from Forbes.com in AI and 5G subsections were collected to build the corpus of Forbes.com in the AI and 5G subsections (A5Fs corpus). The data collected were analyzed and extracted the four-word lexical bundles with the corpus tool known as AntConc (Anthony, 2020).

### 3.1.1 Forbes.com

Forbes is the largest global media brand with numerous international editors, journalists and contributors. The content from these experts is distributed through Forbes.com and Forbes magazine (Forbes Who-We-Are, https://www.forbes.com/connect/who-we-are/). Forbes.com is the online platform of Forbes where contributors, writers or editors all around the world can produce and contribute articles and post them online. Each article is posted under the subsection relating to the content and topic of the article. There are 18 main sections in Forbes.com: Billionaires, Innovation, Leadership, Money, Business, Small Business, Lifestyle, Real Estate, Shopping, Coupons, Advisor, Wheels, Lists, Video, Newsletter, Forbes Magazine, Latest, and Featured. These sections are further broken down into different subsections according to the topic of the article. For example, articles regarding AI, robocars or intelligent machines are posted under AI subsection in the Innovation section, while articles regarding mobile network technology are subsumed under the 5G subsection in the Innovation section.

This study used the A5Fs corpus as the articles were written by trusted contributors, and Forbes' editors review and sometimes edit their articles before posting them online (Steimle, n.d.). The articles written by these experts present authentic and natural language use in journalism, as well as some technical terms in technology and this discipline. Biber and Barbieri (2007) state that lexical bundles serve important discourse functions in both the spoken and written genres. Therefore, this corpus was chosen to analyze the structural and discourse functions of lexical

Ref. code: 25636221042135YPA

bundles used in the articles in Forbes.com. These lexical bundles in A5Fs corpus are believed to represent their natural and authentic use in the journalism register.

Carter and McCarthy (2001) assert that there is no one-size-fits-all corpus, and Bal (2010) indicates that the corpus size must be large enough in order to represent authentic word frequency (Biber, 2006a). However, complete representativeness cannot be achieved in most cases, and the corpus size is determined by capturing an appropriate amount of the language for accurate representation (Reppen, 2010). Representativeness is defined by Biber (1993) as 'the extent to which a sample includes the full range of variability in a population'. In the case of Vaughan (2008), who examined the role of humour in English language teacher faculty meetings at two institutions, a relatively small corpus of 40,000 words was created. In this case, the small corpus is reasonable as it was very specific in context (Reppen, 2010). Flowerdew (2004) also views that the size for a corpus depends on what it contains and what is being investigated. In this study, even though the size of A5Fs corpus was relatively small, the language in this corpus represents the authentic use and is a well-chosen sample of the language in the journalism register.

### 3.2 Data Analysis

Biber and Barbieri (2007) maintain that the four-word lexical items must recur at least 40 times per million words in order to be considered as lexical bundles and the cut-off point is somewhat arbitrary. For Jablonkai (2009), the cut-off point for analyzing four-word lexical bundles in a small corpus is that the four-word lexical bundles should occur at least five times in the corpus in five different texts in order to prevent idiosyncratic language use (Jablonkai, 2009). Therefore, the present study follows the cut-off point introduced by Jablonkai (2009): the four-word lexical bundles should occur at least five times in the corpus in five different texts to be included in the analysis, as the present study contains 447,129 total tokens from 362 texts as shown in Figure 1.

The present study aims to investigate four-word lexical bundles, as Shin and Kim (2017) state that four-word lexical bundles have a positive impact on both high and low proficiency learners' fluency. Four-word bundles are also much more frequent

Ref. code: 25636221042135YPA

than five-word bundles (Hyland, 2008b), “hold three-word bundles in their structures (p. 401)”, and serve a wide variety of functions (Cortes, 2004).

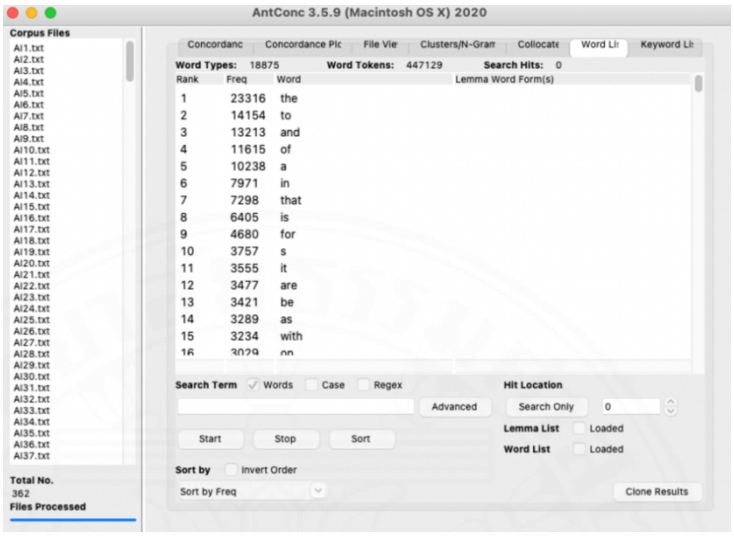



*Figure 1. Total word tokens derived with Word List feature of AntConc (Anthony, 2020)*

AntConc (Anthony, 2020), a concordance software program, was employed to extract the four-word lexical bundles as it is able to identify the frequency of all lexical bundles. Plain text is required to utilize AntConc (Anthony, 2020). Therefore, all 362 articles from Forbes.com were saved as plain text file (.txt).

The data were analyzed in three steps. The first step was to upload all 362 text files into AntConc (Anthony, 2020) by clicking “*Open Dir*” or “*Open File(s)*” in the drop-down of “*File*” menu. The total number of tokens in A5Fs corpus was then identified with the use of Word List function. Then, the “Clusters/N-Grams” feature was used to extract the four-word lexical items. The Min and Max of N-gram Size was set at four and five, respectively, to extract the four-word lexical bundles as the results seem to be deviant when setting the Min and Max of N-gram Size only at four or five. The total number of N-gram tokens was 891,724 as shown in Figure 2, which consisted of the total number of 4-gram tokens (446,043 tokens) and 5-gram tokens (445,681) as shown in Figure 3 and Figure 4, respectively.

Ref. code: 25636221042135YPA

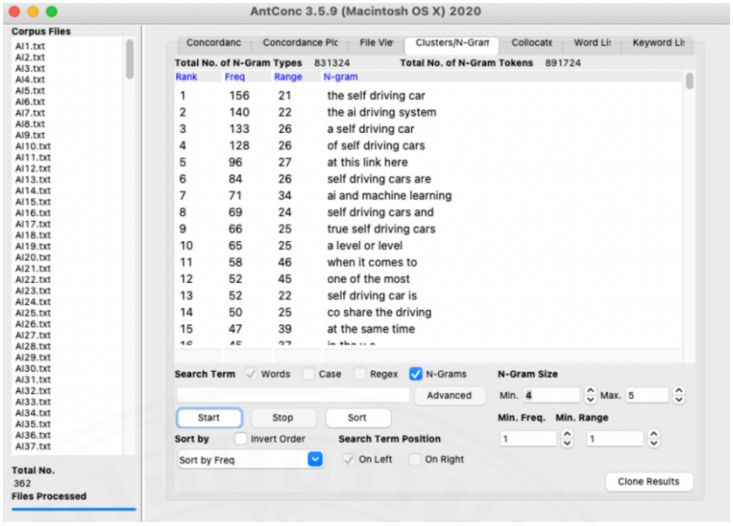



*Figure 2. Data analysis undertaken with the Clusters/N-Grams feature of AntConc (Anthony, 2020)*

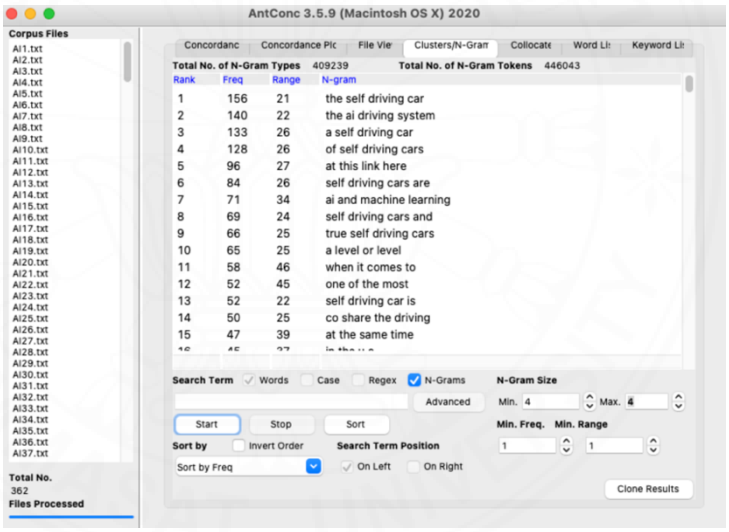



*Figure 3. 4-gram tokens extracted with Clusters/N-Grams feature of AntConc (Anthony, 2020)*

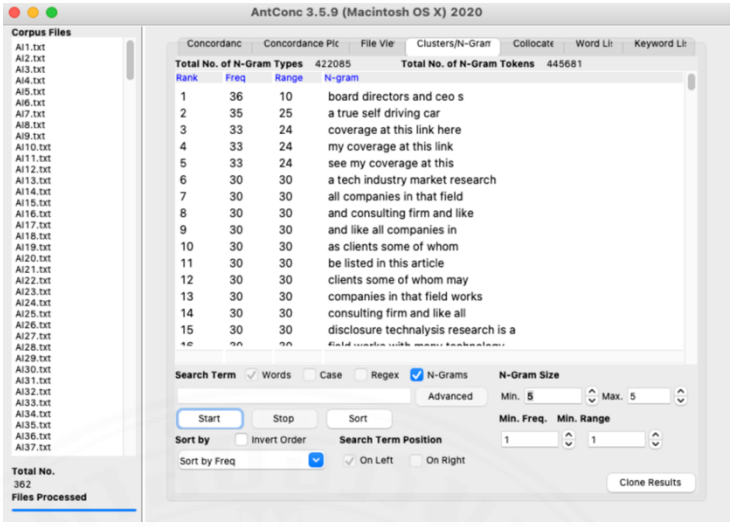



*Figure 4. 5-gram tokens extracted with Clusters/N-Grams feature of AntConc (Anthony, 2020)*

The second step was to manually eliminate the lexical bundles used in the same content, yet in different articles (e.g. *...a tech industry market..., There is not yet..., ...The Level of Self-Driving..., ...the driving actions of..., ...you'll see in a..., ...and there isn't any..., ...car entirely on its..., ...at the wheel of..., ...true self-driving cars...,*) and the lexical bundles with punctuation marks (i.e. *braces, brackets, colon, comma, dash, ellipsis, exclamation mark, parentheses, period, question mark,* and *semicolon*) in between them (e.g. *...a human driver, the..., ...vendors as clients, some...*) from the analysis as the results may be deviant. The lexical bundles which are used in the same content but in different articles were mostly those explaining self-driving cars and the disclosure part from the same writers. Moreover, in order to precisely identify the lexical bundles, orthographic word units, which sometimes combine separate words and contain apostrophes and hyphens (e.g. *it's, isn't, self-driving, co-sharing*), were regarded as one word in the analysis (Biber et al., 2004). However, the present study focused on the top-133 most frequent four-word lexical bundles in A5Fs corpus as time constraints were one of the limitations of this study.

The final step was to analyze the four-word lexical bundles that occurred at least five times in the corpus in five different texts, in accordance with Biber et al.'s (2004) and Hyland's (2008b) taxonomies by investigating the concordance lines of those bundles. The discourse functions of lexical bundles were analyzed and grouped into

Ref. code: 25636221042135YPA

sub-categories according to Biber et al.'s (2004) or Hyland's (2008b) taxonomies. Employing both Biber et al.'s (2004) and Hyland's (2008b) taxonomies in the analysis of the discourse functions of lexical bundles minimized the unproductive comparisons which may arise when grouping bundles in the same sub-categories according to only one taxonomy. Meanwhile, the structural types of lexical bundles were analyzed and grouped into sub-types according to Biber et al.'s (2004) taxonomy. In order to enhance the reliability of the analysis, around 10% of the lexical bundles (15 lexical bundles) were randomly selected to be analyzed by an intercoder using the frameworks employed in this study. The analyses from the intercoder and the researcher were then compared to yield the percentage of agreement, which measures the intercoder reliability. Each bundle was then grouped according to their functions and structures.

### 3.2.1 Intercoder Reliability

There were three steps undertaken to measure the intercoder reliability: intercoder selection, intercoder training, and the analysis of intercoder reliability.

#### 3.2.1.1 Intercoder selection

The intercoder was a student who graduated with a degree of Master of Arts in Career English for International Communication at Thammasat University in the class of 2020. He is now working as an English instructor at Silpakorn University. Although the intercoder is not a native English speaker, his excellent English reading and writing ability qualified him to be the intercoder.

#### 3.2.1.2 Intercoder training

The researcher trained the intercoder to analyze the discourse functions and structures of lexical bundles in accordance with the framework employed in this study. The 10 lexical bundles were presented to the intercoder for the pilot analysis. The intercoder analyzed these bundles independently. Then, the analyses from the researcher and the intercoder were compared. In case there were lexical bundles which were not coded into the same group, the intercoder and the researcher discussed further until the intercoder thoroughly understood the framework and the coding.

#### 3.2.1.3 Analysis of intercoder reliability

The intercoder analyzed 15 randomly selected lexical bundles independently after receiving the training. Once the intercoder analyzed all bundles, the analyses from

Ref. code: 25636221042135YPA

the intercoder and the researcher were compared to measure intercoder reliability, calculated by percent agreement. Lombard et al. (2002) explain that "percent agreement takes value of .00 (no agreement) to 1.00 (perfect agreement)." To calculate the percentage of agreement, the formula presented below was used, in which "A" and "D" are the number of agreements and disagreements, respectively:

$$\frac{A}{(A+D)} \times 100$$

**(1) Intercoder analysis of structural and functional types**

The results of the intercoder structural types and discourse functions analysis are presented in Appendix A and Appendix B, respectively. The intercoder analyzed the 15 randomly selected lexical bundles in the target corpus independently after the training session. The analyses from the intercoder and the researcher were compared to measure intercoder reliability.

The next chapter will present the results of the study on the structural types and functional categories of the lexical bundles in the target corpus. The lists of lexical bundles and the relationship between the structural and functional categories of lexical bundles are also presented and discussed.

Ref. code: 25636221042135YPA

# CHAPTER 4
# RESULTS AND DISCUSSION

## 4.1 Lexical bundles in the A5Fs Corpus

After a thorough analysis of the results from AntConc (Anthony, 2020), the top-133 most frequent four-word lexical bundles in A5Fs corpus are illustrated in Table 4. There were 2,385 total cases from these top-133 most frequent lexical bundles. The most frequent bundles in A5Fs corpus were "*the AI driving system* (5.87%)*, at this link here* (4.03%)*, AI and machine learning* (2.98%)*, when it comes to* (2.43%)*, one of the most* (2.18%)*, at the same time* (1.97%)*, is one of the* (1.64%)*, to be able to* (1.64%)*, the rest of the* (1.59%)*,* and *is going to be* (1.51%)".

**Table 4.** Top-133 most Lexical Bundles in the A5Fs Corpus

| Rank | Frequency | No. of articles | Cluster | Percentage |
|---|---|---|---|---|
| 1 | 140 | 22 | the AI driving system | 5.87 |
| 2 | 96 | 27 | at this link here | 4.03 |
| 3 | 71 | 34 | AI and machine learning | 2.98 |
| 4 | 58 | 46 | when it comes to | 2.43 |
| 5 | 52 | 45 | one of the most | 2.18 |
| 6 | 47 | 39 | at the same time | 1.97 |
| 7 | 39 | 34 | is one of the | 1.64 |
| 8 | 39 | 32 | to be able to | 1.64 |
| 9 | 38 | 20 | the rest of the | 1.59 |
| 10 | 36 | 28 | is going to be | 1.51 |
| 11 | 35 | 31 | in the United States | 1.47 |
| 12 | 34 | 20 | are going to be | 1.43 |
| 13 | 34 | 25 | at the wheel of | 1.43 |
| 14 | 32 | 14 | an AI driving system | 1.34 |
| 15 | 32 | 16 | that the AI driving | 1.34 |
| 16 | 30 | 14 | of the self-driving car | 1.26 |
| 17 | 29 | 27 | as well as the | 1.22 |
| 18 | 28 | 15 | the self-driving car is | 1.17 |
| 19 | 28 | 25 | will be able to | 1.17 |
| 20 | 27 | 11 | the AI driving systems | 1.13 |

Ref. code: 25636221042135YPA

**Table 4.** (Continued)

| Rank | Frequency | No. of articles | Cluster | Percentage |
|---|---|---|---|---|
| 21 | 26 | 22 | a great deal of | 1.09 |
| 22 | 26 | 24 | can be used to | 1.09 |
| 23 | 24 | 12 | self-driving cars will be | 1.01 |
| 24 | 23 | 15 | the odds are that | 0.96 |
| 25 | 22 | 13 | AI driving system is | 0.92 |
| 26 | 22 | 17 | by the end of | 0.92 |
| 27 | 21 | 13 | cars are going to | 0.88 |
| 28 | 21 | 15 | it could be that | 0.88 |
| 29 | 21 | 19 | one of the key | 0.88 |
| 30 | 20 | 18 | in the case of | 0.84 |
| 31 | 20 | 19 | to take advantage of | 0.84 |
| 32 | 19 | 18 | as one of the | 0.80 |
| 33 | 19 | 12 | self-driving cars are going | 0.80 |
| 34 | 18 | 17 | a self-driving car is | 0.75 |
| 35 | 18 | 8 | AI driving systems are | 0.75 |
| 36 | 18 | 18 | for the first time | 0.75 |
| 37 | 18 | 16 | the end of the | 0.75 |
| 38 | 18 | 14 | there will be a | 0.75 |
| 39 | 17 | 17 | at the end of | 0.71 |
| 40 | 16 | 14 | as a result of | 0.67 |
| 41 | 15 | 14 | a wide range of | 0.63 |
| 42 | 15 | 15 | one of the biggest | 0.63 |
| 43 | 15 | 14 | that there is a | 0.63 |
| 44 | 15 | 11 | that there will be | 0.63 |
| 45 | 15 | 7 | the middle of the | 0.63 |
| 46 | 15 | 11 | the point is that | 0.63 |
| 47 | 15 | 12 | the reality is that | 0.63 |
| 48 | 14 | 9 | A to point B | 0.59 |
| 49 | 14 | 14 | are a number of | 0.59 |
| 50 | 14 | 14 | artificial intelligence (AI) and machine | 0.59 |
| 51 | 14 | 9 | from point A to | 0.59 |
| 52 | 14 | 12 | in the form of | 0.59 |
| 53 | 14 | 8 | in the middle of | 0.59 |
| 54 | 14 | 9 | of a self-driving car | 0.59 |

Ref. code: 25636221042135YPA

**Table 4.** (Continued)

| Rank | Frequency | No. of articles | Cluster | Percentage |
|---|---|---|---|---|
| 55 | 14 | 9 | point A to point | 0.59 |
| 56 | 14 | 14 | the future of cars | 0.59 |
| 57 | 13 | 9 | a bit of a | 0.55 |
| 58 | 13 | 9 | are not going to | 0.55 |
| 59 | 13 | 10 | founder and CEO of | 0.55 |
| 60 | 13 | 10 | has the potential to | 0.55 |
| 61 | 13 | 12 | if you want to | 0.55 |
| 62 | 13 | 12 | in the real world | 0.55 |
| 63 | 13 | 13 | keep in mind that | 0.55 |
| 64 | 13 | 9 | of self driving-cars and | 0.55 |
| 65 | 13 | 9 | one of the top | 0.55 |
| 66 | 13 | 12 | that can be used | 0.55 |
| 67 | 13 | 10 | that self-driving cars are | 0.55 |
| 68 | 13 | 13 | there are a number | 0.55 |
| 69 | 12 | 12 | at the heart of | 0.50 |
| 70 | 12 | 9 | at the right time | 0.50 |
| 71 | 12 | 10 | could be that the | 0.50 |
| 72 | 12 | 11 | going to be a | 0.50 |
| 73 | 12 | 9 | has to do with | 0.50 |
| 74 | 12 | 12 | is that they are | 0.50 |
| 75 | 12 | 8 | of AI-based true self | 0.50 |
| 76 | 12 | 11 | of the most important | 0.50 |
| 77 | 12 | 11 | that the self-driving car | 0.50 |
| 78 | 12 | 8 | the last mile of | 0.50 |
| 79 | 12 | 10 | there is also the | 0.50 |
| 80 | 12 | 9 | to cope with the | 0.50 |
| 81 | 11 | 9 | advent of self-driving cars | 0.46 |
| 82 | 11 | 9 | and self-driving tech firms | 0.46 |
| 83 | 11 | 11 | and the ability to | 0.46 |
| 84 | 11 | 9 | automakers and self-driving tech | 0.46 |
| 85 | 11 | 10 | discussion at this link | 0.46 |
| 86 | 11 | 9 | in addition to the | 0.46 |
| 87 | 11 | 8 | in the context of | 0.46 |
| 88 | 11 | 7 | inside a self-driving car | 0.46 |
| 89 | 11 | 10 | my discussion at this | 0.46 |

Ref. code: 25636221042135YPA

**Table 4.** (Continued)

| Rank | Frequency | No. of articles | Cluster | Percentage |
|---|---|---|---|---|
| 90 | 11 | 7 | of self-driving cars is | 0.46 |
| 91 | 11 | 8 | of the AI driving | 0.46 |
| 92 | 11 | 10 | see my discussion at | 0.46 |
| 93 | 11 | 8 | that board directors and | 0.46 |
| 94 | 11 | 9 | the advent of self-driving | 0.46 |
| 95 | 11 | 8 | the use of the | 0.46 |
| 96 | 10 | 7 | AI driving system has | 0.42 |
| 97 | 10 | 10 | as part of a | 0.42 |
| 98 | 10 | 7 | cars on our roadways | 0.42 |
| 99 | 10 | 10 | if there is a | 0.42 |
| 100 | 10 | 10 | in the near future | 0.42 |
| 101 | 10 | 10 | in the world of | 0.42 |
| 102 | 10 | 9 | it is important to | 0.42 |
| 103 | 10 | 9 | said in a statement | 0.42 |
| 104 | 10 | 9 | some of the most | 0.42 |
| 105 | 10 | 9 | take advantage of the | 0.42 |
| 106 | 10 | 5 | the internet of things | 0.42 |
| 107 | 10 | 8 | the notion is that | 0.42 |
| 108 | 10 | 7 | the use of AI | 0.42 |
| 109 | 10 | 7 | use of machine learning | 0.42 |
| 110 | 10 | 6 | use of self-driving cars | 0.42 |
| 111 | 9 | 8 | a self driving car and | 0.38 |
| 112 | 9 | 7 | AI driving system will | 0.38 |
| 113 | 9 | 9 | as part of the | 0.38 |
| 114 | 9 | 8 | as well as a | 0.38 |
| 115 | 9 | 9 | be able to do | 0.38 |
| 116 | 9 | 6 | be able to drive | 0.38 |
| 117 | 9 | 5 | cars on the roadways | 0.38 |
| 118 | 9 | 6 | CEOs and board directors | 0.38 |
| 119 | 9 | 9 | discussed the importance of | 0.38 |
| 120 | 9 | 5 | driving cars and the | 0.38 |
| 121 | 9 | 6 | driving cars on our | 0.38 |
| 122 | 9 | 9 | for a long time | 0.38 |
| 123 | 9 | 6 | in a self-driving car | 0.38 |
| 124 | 9 | 9 | in mind that the | 0.38 |

Ref. code: 25636221042135YPA

**Table 4.** (Continued)

| Rank | Frequency | No. of articles | Cluster | Percentage |
|---|---|---|---|---|
| 125 | 9 | 7 | in the use of | 0.38 |
| 126 | 9 | 8 | in the wake of | 0.38 |
| 127 | 9 | 9 | it is easy to | 0.38 |
| 128 | 9 | 9 | not be able to | 0.38 |
| 129 | 9 | 6 | of AI and machine | 0.38 |
| 130 | 9 | 5 | self-driving cars and the | 0.38 |
| 131 | 9 | 7 | that self-driving cars will | 0.38 |
| 132 | 9 | 9 | the end of this | 0.38 |
| 133 | 9 | 9 | the goal is to | 0.38 |

The most frequent bundles used in the A5Fs corpus, i.e. *the AI driving system*, occured 140 times in 22 different articles (AI15, AI33, AI59, AI69, AI77, AI87, AI93, AI98, AI125, AI129, AI140, AI153, AI162, AI164, AI182, AI190, AI196, AI200, AI210, AI225, AI238, AI248). This bundle provides some information regarding the innovations regarding cars that they are driven or controlled by artificial intelligence (AI) system.

### 4.2 Structural types of lexical bundles

The structural types of lexical bundles were then categorized according to Biber et al.'s (2004) taxonomy presented in Chapter 2. The results showed that of all 133 lexical bundles, 83 (62.41%) lexical bundles incorporate noun phrase and prepositional phrase fragments, 40 (30.07%) lexical bundles incorporate verb phrase fragments, and 10 (7.52%) lexical bundles incorporate dependent clause fragments as shown in Table 5. It can be seen that 62.41% of all lexical bundles were phrasal bundles, while 37.59% were clausal bundles. The writers used all the three structural types, but not all sub-types. However, all sub-types under structural type 3, lexical bundles that incorporate noun phrase and prepositional phrase fragments, were used by the writers. The results also revealed that the most common sub-type was 3d *prepositional phrase expressions,* with a frequency rate of 24.06% (32 bundles) of all lexical bundles. The 3a *(connector+) noun phrase with of-phrase fragment* was the second most common sub-

Ref. code: 25636221042135YPA

type which the writers used in their writings, with a frequency rate of 18.80% (25 bundles) of all lexical bundles. The 1a, 1c, 1f, 1g, and 2a were not used by the writers.

**Table 5.** Structural distribution of lexical bundles in the A5Fs corpus

| **Structural types** | **Number of occurrences** | | | |
|---|---|---|---|---|
| | **Structure sub-types** | **Percentage** | **Token** | **Percentage** |
| **1.Lexical bundles that incorporate *verb phrase* fragments** | **40** | **30.07** | **649** | **27.21** |
| 1a. (connector +) 1st / 2nd person pronoun + VP fragment | 0 | 0.00 | 0 | 0.00 |
| 1b. (connector +) 3rd person pronoun + VP fragment | 19 | 14.29 | 314 | 13.17 |
| 1c. Discourse marker + VP fragment | 0 | 0.00 | 0 | 0.00 |
| 1d. Verb phrase (with non-passive verb) | 20 | 15.04 | 309 | 12.96 |
| 1e. Verb phrase with passive verb | 1 | 0.75 | 26 | 1.09 |
| 1f. *yes-no* question fragments | 0 | 0.00 | 0 | 0.00 |
| 1g. WH-question fragments | 0 | 0.00 | 0 | 0.00 |

**Table 5.** (Continued)

Ref. code: 25636221042135YPA

| Structural types | Number of occurrences | | | |
|---|---|---|---|---|
| | Structure sub-types | Percentage | Token | Percentage |
| **2.Lexical bundles that incorporate *dependent clause* fragments** | **10** | **7.52** | **213** | **8.93** |
| 2a. 1st / 2nd person pronoun + dependent clause fragment | 0 | 0.00 | 0 | 0.00 |
| 2b. WH-clause fragments | 1 | 0.75 | 58 | 2.43 |
| 2c. *If*-clause fragments | 2 | 1.50 | 23 | 0.96 |
| 2d. (verb/adjective +) *to*-clause fragment | 2 | 1.50 | 67 | 2.81 |
| 2e. That-clause fragments | 5 | 3.76 | 65 | 2.73 |
| **3.Lexical bundles that incorporate *noun phrase and prepositional phrase* fragments** | **83** | **62.41** | **1,523** | **63.86** |
| 3a. (connector +) Noun phrase with *of*-phrase fragment | 25 | 18.80 | 392 | 16.44 |
| 3b. Noun phrase with other post-modifier fragment | 4 | 3.01 | 37 | 1.55 |
| 3c. Other noun phrase expressions | 19 | 14.29 | 459 | 19.25 |
| 3d. Prepositional phrase expressions | 32 | 24.06 | 578 | 24.23 |
| 3e. Comparative expressions | 3 | 2.26 | 57 | 2.39 |
| **Total** | **133** | **100.00** | **2,385** | **100.00** |

From Table 6, the most frequent bundle, *the AI driving system,* and the third most frequent bundle, *AI and machine learning*, were under 3c *other noun phrase*

Ref. code: 25636221042135YPA

*expressions,* which was the fourth most common structural sub-type (19 bundles), along with the 1b *(connector +) 3rd person pronoun + VP fragment* structural sub-type with the same number of bundles. Meanwhile, the second and sixth most frequent bundles, *at this link here*, and *at the same time*, were under the most common structural sub-types, 3d *prepositional phrase expressions* (32 bundles). The fifth most frequent bundle, *one of the most*, was under 3a *(connector+) noun phrase with of-phrase fragment*, the second most common structural sub-type (25 bundles). The third common structural sub-type (20 bundles), 1d *verb phrase (with non-passive verb)*, contained the seventh most frequent lexical bundle, *is one of the*.

**Table 6.** Structural analysis of lexical bundles in the A5Fs corpus

| **1** | **Lexical bundles that incorporate *verb phrase* fragments** | |
|---|---|---|
| **1a.** | (connector +) 1st / 2nd person pronoun + VP fragment | - |
| **1b.** | (connector +) 3rd person pronoun + VP fragment | *the self-driving car is, self-driving cars are going, AI driving system has, self-driving cars will be, a self-driving car is, AI driving system will, the odds are that, AI driving systems are, the goal is to, AI driving system is, the point is that, the reality is that, cars are going to, there is also the, it is easy to, it could be that, there are a number, it is important to, there will be a* |
| **1c.** | Discourse marker + VP fragment | - |
| **1d.** | Verb phrase (with non-passive verb) | *is one of the, keep in mind that, said in a statement, is going to be, could be that the, take advantage of the, are going to be, going to be a, see my discussion at, are a number of, has to do with, be able to do, are not going to, is that they are, be able to drive, has the potential to, to cope with the, not be able to, discussed the importance of, to take advantage of* |

**Table 6.** (Continued)

| **1** | **Lexical bundles that incorporate *verb phrase* fragments** |
|---|---|

Ref. code: 25636221042135YPA

| | | |
|---|---|---|
| **1e.** | Verb phrase (with passive verb) | *can be used to* |
| **1f.** | *yes-no* question fragments | - |
| **1g.** | 1g. WH-question fragments | - |
| **2** | **Lexical bundles that incorporate *dependent clause* fragments** | |
| **2a.** | 1st / 2nd person pronoun + dependent clause fragment | - |
| **2b.** | WH-clause fragments | *when it comes to* |
| **2c.** | *If*-clause fragments | *if you want to, if there is a* |
| **2d.** | (verb/adjective +) *to*-clause fragment | *to be able to, will be able to* |
| **2e.** | That-clause fragments | *that there is a, that can be used, that self driving cars will, that there will be, that self-driving cars are* |
| **3** | **Lexical bundles that incorporate *noun phrase and prepositional phrase* fragments** | |
| **3a.** | (connector +) Noun phrase with *of*-phrase fragment | *one of the most, the future of cars, some of the most, the rest of the, founder and CEO of, the internet of things, a great deal of, A bit of a, the use of AI, one of the key, one of the top, use of machine learning, the end of the, the last mile of, use of self-driving cars, as a result of, advent of self-driving cars, as part of the, a wide range of, the advent of self-driving, the end of this. one of the biggest, the use of the, the middle of the, as part of a* |
| **3b.** | Noun phrase with other post-modifier fragment | *the notion is that, driving cars and the, self-driving cars and the, a self-driving car and* |

**Table 6.** (Continued)

Ref. code: 25636221042135YPA

| **3** | **Lexical bundles that incorporate *noun phrase and prepositional phrase* fragments** | |
|---|---|---|
| **3c.** | Other noun phrase expressions | *the AI driving system, and self-driving tech firms, cars on our roadways, AI and machine learning, automakers and self-driving tech, cars on the roadways, an AI driving system, discussion at this link, CEOs and board directors, the AI driving systems, my discussion at this, driving cars on our, artificial intelligence (AI) and machine, and the ability to, point A to point, A to point B, that the AI driving, that the self-driving car, that board directors and* |
| **3d.** | Prepositional phrase expressions | *at this link here, of a self-driving car, of self-driving cars is, at the same time, in the real world, of the AI driving, in the United States, of self-driving cars and, in the near future, at the wheel of, at the heart of, in the world of, of the self-driving car, at the right time, for a long time, by the end of, of AI-based true self, in a self-driving car, in the case of, of the most important, in mind that the, for the first time, in addition to the, in the use of, at the end of, in the context of, in the wake of, in the form of, inside a self-driving car, of AI and machine, in the middle of, from point A to* |
| **3e.** | Comparative expressions | *as well as the, as one of the, as well as a* |

### 4.3 Discourse functions of lexical bundles

The discourse functions of lexical bundles in the A5Fs corpus were analyzed and categorized according to Biber et al.'s (2004) and Hyland's (2008b) taxonomies. The analysis incorporated the two taxonomies to minimize the comparisons when grouping lexical bundles under the same sub-categories employing one taxonomy. Moreover, the sub-categories under one taxonomy may not cover all discourse functions of the lexical bundles in the target corpus. However, some sub-categories from the two taxonomies shared lexical bundles with the target corpus. As can be seen from Table 7, 8, and 9, the lexical bundles under the quantity sub-category in Biber et al.'s (2004) taxonomy were shared with the quantification sub-category in Hyland's (2008b) taxonomy as they both describe amount or number. As the location sub-

Ref. code: 25636221042135YPA

category in Hyland's (2008b) taxonomy indicates time or place, the combination of lexical bundles under place, time, and multifunctional reference sub-categories in Biber et al.'s (2004) taxonomy were grouped under the location sub-category used by Hyland (2008b). The stance bundles category and stance features sub-categories of the two taxonomies can also share lexical bundles conveying the writers' attitude, judgement, certainty or uncertainty. Therefore, a total of 56 lexical bundles were shared between the two taxonomies, which contained 942 cases. There were 28, 22, and six lexical bundles from the total of 56 shared lexical bundles under the stance features, location, and quantification sub-categories in Hyland's (2008b) taxonomy, respectively.

The results showed that Hyland's (2008b) research-oriented bundles were the most common category in the target corpus, with the frequency rate of 40.60% (54 bundles). The second most common category was Biber et al.'s (2004) referential bundles, with the frequency rate of 37.59% (50 bundles). Both research-oriented and referential bundles contained sub-categories which shared lexical bundles indicating time, location, and quantity. The third and fourth most common categories were Hyland's (2008b) participant-oriented (31 bundles), and Biber et al.'s (2004) stance bundles (28 bundles). The lexical bundles under these two categories were shared as they convey the writers' attitude, evaluation, certainty or uncertainty. There were seven sub-categories with no occurrences in the target corpus: epistemic stance, intangible framing, imprecision, politeness, simple inquiry, transition signals, and structuring signals from two taxonomies.

**Table 7.** Discourse functions distribution of lexical bundles in the A5Fs corpus

| **Discourse functions categories (Biber et al., 2004)** | **Number of occurrences** | | | |
|---|---|---|---|---|
| | **Sub-categories** | **Percentage** | **Token** | **Percentage** |
| **1. Stance bundles** | **28** | **21.05** | **470** | **19.71** |
| 1.1 Epistemic stance | 0 | 0.00 | 0 | 0.00 |
| 1.2 Attitudinal/modality stance | | | | |
| Desire | 1 | 0.75 | 13 | 0.55 |
| Obligation/directive | 4 | 3.01 | 51 | 2.14 |
| Intention/prediction | 15 | 11.28 | 262 | 10.99 |

**Table 7.** (Continued)

| | **Number of occurrences** |
|---|---|

Ref. code: 25636221042135YPA

| **Discourse functions categories (Biber et al., 2004)** | **Sub-categories** | **Percentage** | **Token** | **Percentage** |
|---|---|---|---|---|
| 1.2 Attitudinal/modality stance | | | | |
| Ability | 8 | 6.02 | 144 | 6.04 |
| **2. Discourse organizers** | **20** | **15.04** | **367** | **15.39** |
| 2.1 Topic introduction | 2 | 1.50 | 23 | 0.96 |
| 2.2 Topic elaboration/clarification | 18 | 13.53 | 344 | 14.42 |
| **3. Referential expressions** | **50** | **37.59** | **926** | **38.83** |
| 3.1 Specification of attributes | | | | |
| Intangible framing | 0 | 0.00 | 0 | 0.00 |
| Tangible framing | 3 | 2.26 | 38 | 1.59 |
| Quantity | 6 | 4.51 | 114 | 4.78 |
| 3.2 Identification/focus | 17 | 12.78 | 389 | 16.31 |
| 3.3 Imprecision | 0 | 0.00 | 0 | 0.00 |
| 3.4 Time/place/text reference | | | | |
| Place | 12 | 9.02 | 182 | 7.63 |
| Time | 2 | 1.50 | 59 | 2.47 |
| Text deixis | 2 | 1.50 | 27 | 1.13 |
| Multifuctional reference | 8 | 6.02 | 117 | 4.91 |
| **4. Special conversational functions** | **1** | **0.75** | **10** | **0.42** |
| 4.1 Politeness | 0 | 0.00 | 0 | 0.00 |
| 4.2 Simple inquiry | 0 | 0.00 | 0 | 0.00 |
| 4.3 Reporting | 1 | 0.75 | 10 | 0.42 |
| **Discourse functions categories (Hyland, 2008b)** | | | | |
| **1. Research-oriented** | **54** | **40.60** | **975** | **40.88** |
| 1.1 Location | 22 | 16.54 | 358 | 15.01 |
| 1.2 Procedure | 9 | 6.77 | 90 | 3.77 |
| 1.3 Quantification | 6 | 4.51 | 114 | 4.78 |
| 1.4 Description | 4 | 3.01 | 48 | 2.01 |
| 1.5 Topic | 13 | 9.77 | 365 | 15.30 |

**Table 7.** (Continued)

Ref. code: 25636221042135YPA

| Discourse functions categories (Hyland, 2008b) | Number of occurrences | | | |
|---|---|---|---|---|
| | Sub-categories | Percentage | Token | Percentage |
| **2. Text-oriented** | **5** | **3.76** | **66** | **2.77** |
| 2.1 Transition signals | 0 | 0.00 | 0 | 0.00 |
| 2.2 Resultative signals | 1 | 0.75 | 16 | 0.67 |
| 2.3 Structuring signals | 0 | 0.00 | 0 | 0.00 |
| 2.4 Framing signals | 4 | 3.01 | 50 | 2.10 |
| **3. Participant-oriented** | **31** | **23.31** | **513** | **21.51** |
| 3.1 Stance | 28 | 21.05 | 470 | 19.71 |
| 3.2 Engagement | 3 | 2.26 | 43 | 1.80 |
| **Shared lexical bundles** | **56** | **42.11** | **942** | **39.50** |
| **Total** | **133** | **100.00** | **2,385** | **100.00** |

**Table 8.** Discourse functions analysis of lexical bundles in the A5Fs corpus according to Biber et al.'s (2004) taxonomy

| Categories | Sub-categories | Bundles from the A5Fs corpus |
|---|---|---|
| **1. Stance bundles** | 1.1 Epistemic stance | - |
| | 1.2 Attitudinal/modality stance | |
| | Desire | *if you want to* |
| | Obligation/directive | *has to do with, it is important to, to take advantage of, in mind that the* |
| | Intention/prediction | *AI driving system will, if there is a, self-driving cars will be, are going to be, is going to be, that self-driving cars will, are not going to, it could be that, that there will be, cars are going to, it is easy to, there will be a, going to be a, self-driving cars are going, could be that the* |
| | Ability | *and the ability to, can be used to, to be able to, be able to do, has the potential to, will be able to, be able to drive, not be able to* |

**Table 8.** (Continued)

| Categories | Sub-categories | Bundles from the A5Fs corpus |
|---|---|---|
| **2. Discourse organizer** | 2.1 Topic introduction | *in addition to the, there is also the* |
| | 2.2 Topic elaboration/ clarification | *a self-driving car is, is that they are, that the self-driving car, AI driving systems are, of self-driving cars is, that there is a, AI driving system has, that board directors and, the notion is that, AI driving system is, that can be used, the odds are that, as well as a, that self-driving cars are, the self-driving car is, as well as the, that the AI driving, when it comes to* |
| **3. Referential expressions** | 3.1 Specification of attributes | |
| | Intangible framing | - |
| | Tangible framing | *a wide range of, as part of the, in the form of* |
| | Quantity | *a bit of a, are a number of, the rest of the, a great deal of, some of the most, there are a number* |
| | 3.2 Identification/focus | *as one of the, of AI and machine, one of the biggest, at this link here, of AI-based true self, one of the key, discussion at this link, of self-driving cars and, one of the most, is one of the, of the AI driving, one of the top, my discussion at this, of the most important, see my discussion at, of a self-driving car, of the self driving car* |
| | 3.3 Imprecision | - |
| | 3.4 Time/place/text reference | |
| | Place | *A to point B, driving cars on our, in the real world, at the wheel of, from point A to, in the United States, cars on our roadways, in a self-driving car, in the world of, cars on the roadways, inside a self-driving car, point A to point* |
| | Time | *at the right time, at the same time* |
| | Text deixis | *for a long time, for the first time* |
| | Multifuctional reference | *at the end of, in the middle of, the end of this, at the heart of, in the near future, the middle of the, by the end of, the end of the* |

**Table 8.** (Continued)

Ref. code: 25636221042135YPA

| Categories | Sub-categories | Bundles from the A5Fs corpus |
|---|---|---|
| **4. Special conversational functions** | 4.1 Politeness | - |
| | 4.2 Simple inquiry | - |
| | 4.3 Reporting | *said in a statement* |

**Table 9.** Discourse functions analysis of lexical bundles in the A5Fs corpus according to Hyland's (2008b) taxonomy

| Categories | Sub-categories | Bundles from the A5Fs corpus |
|---|---|---|
| **1. Research-oriented** | 1.1 Location | *A to point B, cars on the roadways, in the real world, at the end of, driving cars on our, in the United States, at the heart of, from point A to, in the world of, at the right time, in a self-driving car, point A to point, at the same time, inside a self-driving car, the end of the, at the wheel of, in the middle of, the end of this, by the end of, in the near future, the middle of the, cars on our roadways* |
| | 1.2 Procedure | *discussed the importance of, to cope with the, the use of the, in the use of, the goal is to, use of machine learning, take advantage of the, the use of AI, use of self-driving cars* |
| | 1.3 Quantification | *a bit of a, are a number of, the rest of the, a great deal of, some of the most, there are a number* |
| | 1.4 Description | *advent of self-driving cars, the future of cars, the last mile of, the advent of self-driving* |
| | 1.5 Topic | a self-driving car and, automakers and self-driving tech, self-driving cars and the, AI and machine learning, ceos and board directors, the AI driving system, an AI driving system, driving cars and the, the AI driving systems, and self-driving tech firms, founder and CEO of, the internet of things, artificial intelligence (AI) and machine |

**Table 9.** (Continued)

Ref. code: 25636221042135YPA

| Categories | Sub-categories | Bundles from the A5Fs corpus |
|---|---|---|
| **2. Text-oriented** | 2.1 Transition signals | - |
| | 2.2 Resultative signals | *as a result of* |
| | 2.3 Structuring signals | - |
| | 2.4 Framing signals | *as part of a, in the context of, in the wake of, in the case of* |
| **3. Participant-oriented** | 3.1 Stance features | *AI driving system will, has to do with, not be able to, and the ability to, has the potential to, self-driving cars are going, are going to be, if you want to, self-driving cars will be, are not going to, if there is a, to be able to, be able to do, in mind that the, to take advantage of, be able to drive, is going to be, that self-driving cars will, can be used to, it could be that, that there will be, cars are going to, it is easy to, there will be a, could be that the, it is important to, will be able to, going to be a* |
| | 3.2 Engagement features | *keep in mind that, the point is that, the reality is that* |

### 4.3.1 Biber et al.'s (2004) discourse functions analysis

#### 4.3.1.1 Stance bundles

Writers convey their attitudes, judgement, ability, certainty or uncertainty through stance bundles. As presented in Chapter 2, stance bundles consist of two main sub-categories: epistemic stance and attitudinal/modality stance (Biber et al., 2004). The results showed that this category was the third most common category, with the frequency rate of 21.05% (28 bundles). However, there was no bundle belonging to this epistemic stance sub-category which provides information on certainty, uncertainty or possibility.

Writers deliver their attitudes towards actions or events through the attitudinal/modality stance. In this study, stance bundles were the fourth most common categories employed by the writers. Intention/prediction sub-category had the highest frequency rate at 11.28% (15 bundles) under this category. As illustrated in Table 8, all

Ref. code: 25636221042135YPA

the lexical bundles under this sub-category convey the message of prediction rather than intention. This may be because the writers tried to predict technological innovations in the near future. The three most frequent bundles under this sub-category were "*is going to be* (36 times), *are going to be* (34 times), and *self-driving cars will be* (24 times)". The examples are presented below.

"As with many things 5G, the process of densifying and improving the performance of 5G networks ***is going to be*** a relatively long, complicated road."

(*'is going to be'*, 5G-7)

"Eventually, on a competitive basis, they will all need to pretty much meet the same desired capabilities, else those self-driving cars not so outfitted ***are going to be*** avoided by the paying public."

(*'are going to be'*, AI196)

"In however manner it comes to be, let's assume that ***self-driving cars will be*** principally used on a ridesharing basis."

(*'self-driving cars will be'*, AI179)

The ability sub-category was the second most common sub-category under stance bundles, with the frequency rate of 6.02% (eight bundles). These bundles indicate the ability of the subject mentioned in the article. The three most frequent bundles used under this sub-category were "*to be able to* (39 times), *will be able to* (28 times), and *can be used to* (26 times)". The examples are illustrated below.

"A battery-powered one-person aircraft that we're unlikely ***to be able to*** buy any time very soon, but gives an exciting vision of how personal transport is evolving."

(*'to be able to'*, AI9)

"In particular, these new signal types ***can be used to*** notify devices that they can temporarily put their modems, RF front ends, and other components in low power mode to preserve energy."

(*'can be used to'*, 5G-15)

"New opportunities to drive revenue, increase innovation, and connect with people will abound. For example, simple mobile devices ***will be able to*** do more with

Ref. code: 25636221042135YPA

virtual reality and augmented reality, which could in itself bring new business prospects."

(*'will be able to'*, 5G-44)

The lexical bundles under the obligation/directive and desire sub-categories were the least frequent occurred lexical bundles in the attitudinal/modality stance, with frequency rates of 3.01% (four bundles) and 0.75% (one bundle), respectively. The obligation/directive bundles convey the message that something needs to be done in a particular way, while the desire bundles indicate the writers' or someone's needs. The examples of obligation/directive and desire bundles are presented below, respectively.

"Conversely, even if a network offers the potential for newer features, faster speeds, etc., an individual device won't be able ***to take advantage of*** them unless it has a modem that supports those functions."

('*to take advantage of*', 5G-98)

"So, for example, ***if you want to*** separate time-sensitive IoT traffic for manufacturing-based sensors from the data traffic on a consumer's smartphone or even provide a private VPN-style connection for a company's employees, network slicing can provide this and much, much more."

('*if you want to'*, 5G-41)

#### 4.3.1.2 Discourse organizers

Discourse organizers are used to introduce, clarify or elaborate on a topic both before and after discourse. This category was the fifth most common, with the frequency rate of 15.04% (20 bundles). There were 18 bundles used to provide more information about the topic discussed or a previous statement in the article (e.g. *when it comes to* (58 times), *that the AI driving* (32 times)), while two bundles were used to introduce a new topic (e.g. *there is also the* (12 times)). The examples are presented below.

"The Snapdragon 888's GPU is a beast ***when it comes to*** performance, thanks to the new Adreno 660 architecture."

*('when it comes to'*, 5G-77)

Ref. code: 25636221042135YPA

"In any case, the side point here is ***that the AI driving*** systems and the Natural Language Processing (NLP) interactivity with passengers has to get further attention and inevitably will be a significant determiner of how people will choose the brand of a self-driving car to use."

(*'that the AI driving'*, AI248)

"Likewise, the end of the pandemic could shift consumer spending to other areas, especially travel. ***There is also the*** concern that prior to the pandemic, most of the global economy was in a prolonged growth period that has been minimally affected by the pandemic."

(*'there is also the'*, 5G-83)

#### 4.3.1.3 Referential expressions

There were a total of 50 lexical bundles (37.59%) under referential expressions. The only sub-category with no lexical bundles was imprecision. There were six lexical bundles used to describe the amount or number (e.g. *the rest of the* (38 times), *a great deal of* (26 times)), and three bundles specifying tangible framing (e.g. *a wide range of* (15 times)). There were 17 lexical bundles used to indicate the focus of something discussed or mentioned in the text (e.g. *at this link here* (96 times), *one of the most* (52 times), and *is one of the* (39 times)). The examples are presented below.

"Turning to the other commonly experienced situation, consider a scenario involving an AV unit that is cloud immersed, and meanwhile, ***the rest of the*** company has yet to make that same shift."

(*'the rest of the'*, AI108)

"The bottom-line takeaway from all this data is that there are definitely situations in which deploying mmWave as a supplement to existing mid-band based 5G services makes ***a great deal of*** sense and other situations for which the timing could take a bit longer."

(*'a great deal of'*, 5G-91)

Ref. code: 25636221042135YPA

"The system presents ***a wide range of*** content for each bottle, educating owners as their collection grows."

(*'a wide range of'*, AI229)

"You know that the AI is a highly safe driver, especially in comparison to a human driver (this has yet to be shown, and we might end-up with AI driving systems that are only as good as human drivers in terms of safety, perhaps worse, see my explanation ***at this link here***)."

(*'at this link here'*, AI7)

"By using the drone, the AI can figure out what might be a prudent path to get to the desired destination (for more about drones and self-driving cars, see my analysis ***at this link here***).

(*'at this link here'*, AI77)

"Qualcomm has successfully led the 4G and 5G charge, and its lineup offers ***one of the most*** diverse selection of modems available."

(*'one of the most'*, 5G-105)

"What has not received much attention, however, is the industries that are growing and adding jobs during these troubling times. Not surprisingly, the tech industry ***is one of the*** fortunate few that's doing well."

(*'is one of the'*, 5G-48)

There were a total of 24 lexical bundles indicating place, time, or text reference. From the total of 24 bundles, 12 bundles indicated place (e.g. *in the United States* (35 times), *at the wheel of* (34 times)), eight bundles can be used multifunctionally to indicate both place and time (e.g. *the end of the* (18 times)), two bundles indicated time, and two bundles were used as text deixis. The examples are presented below.

"Founded in 1938, over 80 years ago, TSC has grown into the largest farm store operator ***in the United States***."

(*'in the United States'*, 5G-96)

Ref. code: 25636221042135YPA

"Any and all driving robots that are going to be ***at the wheel of*** a car must be reliable, safe, secure, trustworthy, etc., and imbue a slew of such vital and absolutely required traits."

(*'at the wheel of'*, AI93)

"The odds are that you would work in a tall skyscraper. Up the elevator each morning, and then down you come at ***the end of the*** workday."

(*'the end of the'*, AI179)

#### 4.3.1.4 Special conversational functions

There was only one lexical bundle used to introduce a reporting clause, i.e. *said in a statement* (10 times). There were no lexical bundles used to introduce politeness or simple inquiry. The instance is presented below.

"As a result, money from the FCC's $8.3 billion Universal Service Fund can no longer be used to "purchase, obtain, maintain, improve, modify, or otherwise support" any equipment or services produced or provided by Huawei and ZTE, the agency ***said in a statement*** today."

(*'said in a statement'*, 5G-9)

### 4.3.2 Hyland's (2008b) discourse functions analysis

#### 4.3.2.1 Research-oriented

Research-oriented bundles indicate or explain locations, procedures, quantifications, and give descriptions of research objects and topics related to the research (Hyland, 2008a). As can be seen from Table 7, there were a total of 54 lexical bundles (40.60%) under this category. The location and quantification sub-categories consisted of 22 shared lexical bundles under place, time, and multifunctional reference and six shared lexical bundles under the quantity sub-categories in Biber et al.'s (2004) taxonomy. As illustrated in Table 9, there were 13 lexical bundles related to the topic of the article (e.g. *the AI driving system* (140 times), *AI and machine learning* (71 times)). There were nine lexical bundles used to indicate the methodology or purpose of the statement (e.g. *the use of the* (11 times), *the goal is to* (nine times)). The other

Ref. code: 25636221042135YPA

four lexical bundles were used to provide the qualities or properties of materials mentioned in the article (e.g. *the future of cars* (14 times)). The examples are presented below.

"In the case of scavenger hunts, you can readily argue that it is fantastic that ***the AI driving systems*** won't drive illegally."

(*'the AI driving systems'*, AI225)

"Another area that Qualcomm is keenly focused on is ***AI and machine learning***, which are technologies that modern mobile platforms utilize in almost every facet of the user experience, from photography to recommendation engines, to gaming and even power efficiency."

(*'AI and machine learning'*, 5G-76)

"Besides the raw computing power, there is also ***the use of the*** sensory devices that will be on autonomous vehicles (and, to some degree, also included on semi-autonomous too)."

(*'the use of the'*, AI220)

"***The goal is to*** allow network equipment vendors and carriers to be able to build a variety of different open standards-based 5G network-optimized infrastructure components."

(*'the goal is to'*, 5G-60)

"Speaking of cars, consider ***the future of cars*** and the advent of self-driving cars."

(*'the future of cars'*, AI248)

#### 4.3.2.2 Text-oriented

Text-oriented bundles are used to provide ways to engage with the article, provide warrants, connect ideas, direct readers around the text, and specify limitations (Hyland, 2008b). As can be seen from Table 9, there were five lexical bundles under this category. There are four lexical bundles under framing signals which situate the statement by specifying the limitations (e.g. *in the context of* (11 times)), and one lexical bundle under resultative signals, which indicates a causative relationship between elements (e.g. *as a result of* (16 times)). *'In the context of'* in the example frames that

Ref. code: 25636221042135YPA

the statement is about aviation, while *'as a result of'* in the example shows that the following statement is the cause of the previous statement.

"In homes, on roads, and elsewhere, there will be multiple vendors. Each group of robots will have different goals and different communications, making interaction difficult. The authors then bring in the context of aviation. The FAA created an accident database which all industry partners use to input data and perform analysis. (AI12)"

(*'in the context of'*, AI12)

"The technology wasn't as glamorous as big iron processors or high speed communications chips, but FPGAs played pivotal a role in a huge number of systems designs over the years, ***as a result of*** their intrinsic time-to-market advantages and flexibility."

(*'as a result of'*, 5G-64)

#### 4.3.2.3 Participant-oriented

As can be seen from Table 7, there were 31 lexical bundles (23.31%) under this category. There were 28 lexical bundles under stance features sub-category used to convey the writers' attitude and evaluation, which correspond to the stance bundles in Biber et al.'s (2004) taxonomy. The other three lexical bundles were used to address readers (e.g. *the point is that* (15 times), *keep in mind that* (13 times)). The examples are presented below.

"The driving robot must be able to drive the car entirely on its own. In this ground rule, ***the point is that*** the driving robot cannot "cheat" and somehow require someone or something else to aid it in doing the driving."

(*'the point is that'*, AI93)

"Investors planning a joint venture or other investment activities should also ***keep in mind that*** CFIUS has jurisdiction to review any foreign investment, even retroactively cleared transactions."

(*'keep in mind that'*, 5G-50)

Ref. code: 25636221042135YPA

## 4.4 Relationship between structural and functional categories of lexical bundles in the A5Fs corpus

Biber et al. (2004) state that "most referential bundles are composed of noun phrase or prepositional phrase fragments". While according to Hyland (2008b), research-oriented bundles have the *(connector +) NP with of-phrase* structure. The results from this study revealed that most phrased-based bundles were referential and research-oriented bundles. There were 46 (92%) phrase-based bundles from the total of 50 referential bundles as shown in Figure 5. From the total of 54 research-oriented bundles, there were 48 (88.89%) phrase-based bundles. *'A great deal of, the end of the,* and *some of the most'* are the examples of the lexical bundles with *(connector +) NP with of-phrase* structure which were categorized under referential and research-oriented bundles. Meanwhile, most stance bundles and participant-oriented bundles were composed of clause-based bundles. There were 26 (92.86%) clause-based bundles from the total of 28 stance bundles, and 29 (93.55%) clause-based bundles from the total of 31 participant-oriented bundles. The relationship between the structural and functional categories of lexical bundles in this study mirrors the studies investigating journalistic discourse (Dastjerdi & Rafiee, 2013; Fadhilah, 2020). Dastjerdi and Rafiee (2013) found that most phrasal fragments form referential bundles, and clausal fragments form stance bundles.

Ref. code: 25636221042135YPA

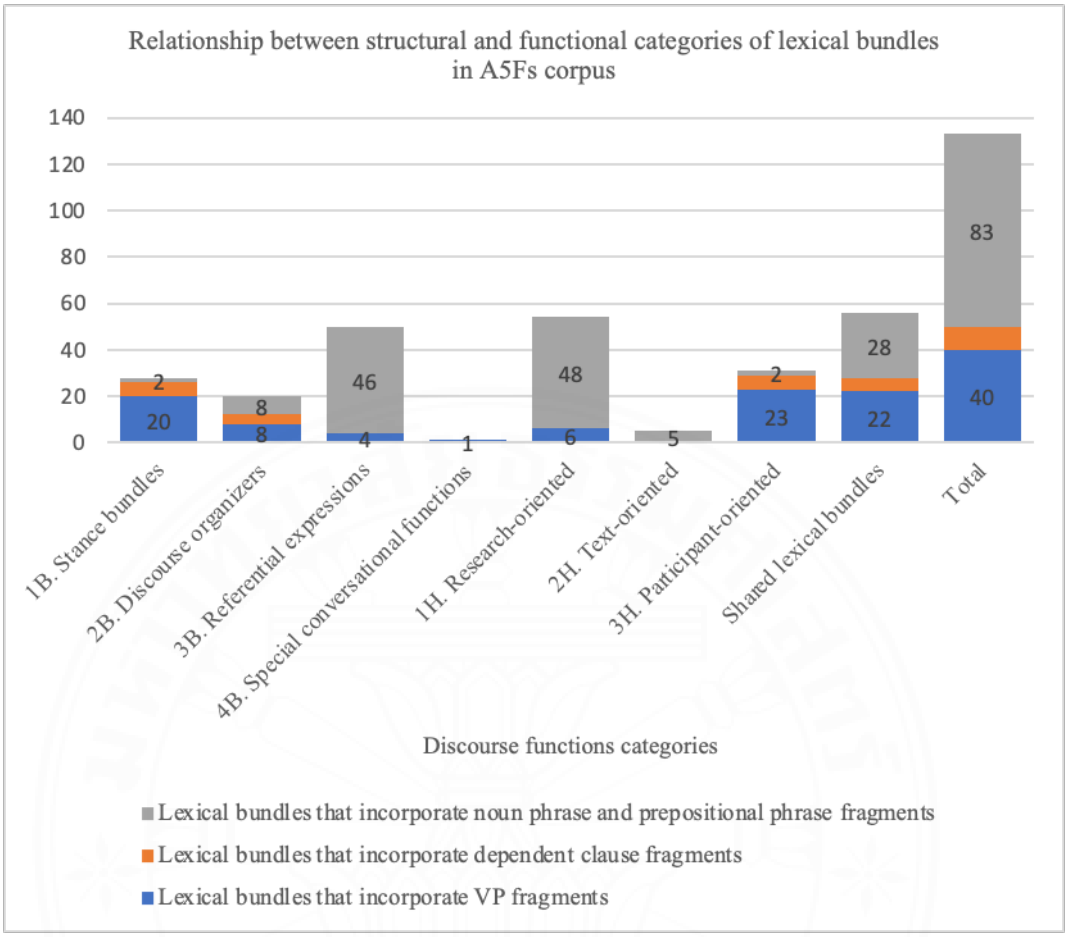



*Figure 5. Relationship between structural and functional categories of the A5Fs corpus*

## 4.5 Discussion

The lexical bundles in this study were collected from the written articles in Forbes.com. Therefore, the results from this study and other studies analyzing the structural and functional types of lexical bundles in the journalistic register are in agreement with several studies focusing on the academic written register. The relationship between the structural and functional types from this study also correspond with other studies (Biber et al., 2004; Hyland, 2008b; Dastjerdi & Rafiee, 2013; Fadhilah, 2020). It is noteworthy that the use of lexical bundles in the journalistic register is in line with the academic discourse.

The structural and functional analyses of lexical bundles in A5Fs corpus conform to the results of several studies focusing on academic discourse (Cortes, 2004;

Ref. code: 25636221042135YPA

Hyland, 2008b; Juknevičien, 2009; Pang, 2009; Chen & Baker, 2010; Llanta, 2014; Qin, 2014; Pan et al., 2015; Güngör & Uysal, 2016) and the journalism register (Rafiee et al., 2011; Dastjerdi & Rafiee, 2013; Fadhilah, 2020). The results from the studies focusing on academic discourse revealed that native English writers used more bundles consisting of noun and prepositional phrases than non-native English writers did (Cortes, 2004; Hyland, 2008b; Juknevičien, 2009; Pang, 2009; Chen & Baker, 2010; Llanta, 2014; Qin, 2014; Pan et al., 2015; Güngör & Uysal, 2016). The results from the studies focusing on journalism register also revealed that the journalists used more phrase-based bundles than clause-based bundles (Rafiee et al., 2011; Dastjerdi & Rafiee, 2013). The results from this study showed that the writers of Forbes.com employed more phrase-based bundles (83 bundles) than clause-based bundles (50 bundles).

The most frequent discourse function of lexical bundles in this study was Hyland's (2008b) research-oriented function, which corresponds with Biber et al.'s (2004) referential bundles. The results of the functional analysis of the lexical bundles in this study were also in line with several studies focusing on academic discourse and the journalism register. The results from the studies in academic and journalistic discourse indicated that native English writers relied heavily on referential or research-oriented bundles (Chen & Baker, 2010; Güngör & Uysal, 2016; Pan et al., 2016), and the referential bundles are more frequent than the other categories (Dastjerdi & Rafiee, 2013; Rafiee & Keihaniyan, 2013; Fadhilah, 2020).

Ref. code: 25636221042135YPA

# CHAPTER 5
# CONCLUSION AND RECOMMENDATIONS

## 5.1 Conclusion

The results from this study shed some light on the lexical bundles used in online technology innovation articles. The results should be beneficial for English for specific purpose teachers, journalism students, novice journalists, and journalists who wish to enhance their writing skills.

The results revealed that both structures and discourse functions of lexical bundles in the journalism register are aligned with the academic discourse (Cortes, 2004; Hyland, 2008b; Juknevičien, 2009; Pang, 2009; Chen & Baker, 2010; Llanta, 2014; Qin, 2014; Pan et al., 2015; Güngör & Uysal, 2016). Therefore, this implies that the lexical bundles employed in the academic discourse and journalism register can be used interchangeably, although there were some specific lexical bundles used in specific registers (e.g. *the AI driving system*, *AI and machine learning*). The relationship between the structural and functional types from this study also showed that the referential or research-oriented bundles as well as the stance or participant-oriented bundles were comprised of phrase-based, and clause-based bundles, respectively, which corresponds with other studies (Biber et al., 2004; Hyland, 2008b; Dastjerdi & Rafiee, 2013; Fadhilah, 2020).

## 5.2 Pedagogical implications

English for specific purpose teachers can apply the results to benefit their students who aim to work in this field. It is important to ensure that learners are aware of the use of lexical bundles so that they can apply them in their utterances much more effectively (Lewis, 1997). Therefore, appropriate material design should be applied (Lewis, 1997). In order to enhance learners' understanding of the use of lexical bundles, they should also be contextualized through the text (Chun-guang, 2014). With effective and appropriate course and material design incorporating the results from this study,

Ref. code: 25636221042135YPA

not only will students' writing skills be enhanced but also their confidence in their writing (Yoon & Hirvela, 2004).

Both learners and teachers can use the results from this corpus as a pedagogical resource, tool, and database (McCarthy, 1999; Yoon & Hirvela, 2004; O'Keefe et al., 2007). Several previous studies have indicated that corpora can be both pedagogical resources and tools through which learners can explore language use in their areas of interest (McCarthy, 1999; Yoon & Hirvela, 2004). Even though there are advantages to incorporating a corpus-driven approach in language teaching, few studies have reported significant improvement in learners' reading and writing competencies (Chen, 2008). However, learners perceived that corpus can be applied to help enhance their usage of phrases and vocabulary, writing skills and increase their confidence in writing (Yoon & Hirvela, 2004).

Hence, the results of this study can be used in the stage of material design, course design or course development. Some lexical bundles displayed the trends and inventions related to technology. Others provided the insights into online article writing practices. As Bareggi (2006) recommended, students' awareness of lexical "chunking" should be raised through language teaching, while their ability to "chunk" language successfully should also be developed through language teaching (Bareggi, 2006). The teacher may assign the students to write an article about technology related to their interests. Then, the teacher can use a concordance tool, such as AntConc (Anthony, 2020), to retrieve the lexical bundles used in the students' writings and compare them with the lexical bundles in the target corpus, followed by discussion and explanations of the structures and functions in the discourse. The teachers can also assign the students to analyze their own lexical bundles employed in their writings so that their awareness is increased and their writing skills are enhanced.

Ref. code: 25636221042135YPA

## 5.3 Limitations and Recommendations

The present study provides some useful insights into lexical bundles in the target corpus which can be applied in journalism writing. The limitations of this study and some recommendations for further studies are presented below:

5.3.1 The total words in this corpus were 447,129 from 362 texts, which yielded almost 500,000 lexical bundles. Due to time limitations, only the top-133 most frequent lexical bundles were analyzed. In light of this, further study can explore and analyze the structures and discourse functions of lexical bundles in this corpus.

5.3.2 This study analyzed only the articles regarding technological innovations in Forbes.com. Further research can be expanded to other sources of media (e.g. magazines, news article, news report, etc.) to shed more light on the lexical bundles in the journalism register both in the spoken and written genres.

5.3.3 This study also only focused on the lexical bundles in the target corpus. Therefore, some interesting jargon was not taken into account. Further investigations of jargon in technology and innovation articles can be done with the focus on their collocates for better understanding and usage by journalists or journalism students.

5.3.4 The collected articles from Forbes.com were written by both native and non-native English writers. However, this study did not conduct contrastive analysis of articles written by native and non-native English writers. Further study can compare articles by these types of writers, as the results may reveal some interesting insights.

5.3.5 Articles on Forbes.com are posted almost every day, and there may be new information or trends that emerge. Moreover, as there are new inventions and discoveries relating to technology daily, the articles collected in this study may be obsolete shortly. Therefore, future research may need to collect new articles for further study.

5.3.6 The percent agreement of the discourse functions analyses between the researcher and the intercoder was 0.73 (73%), failing to reach the minimum acceptable interrater agreement of (McHugh, 2012). Due to some ambiguity in the discourse functions in the lexical bundles, the researcher and intercoder might have analyzed and grouped the bundles differently. Further study may inquire about the judgements or

Ref. code: 25636221042135YPA

opinions of both the researcher and intercoder to shed light on their decisions to group or analyze the discourse functions of each lexical bundle into specific categories.

Ref. code: 25636221042135YPA

# REFERENCES

Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Bal, B. (2010). *Analysis of four-word lexical bundles in published research articles written by Turkish scholar.* [Unpublished master thesis]. Georgia State University.

*Bangkok Post*. (2019). Bangkokpost. Retrieved February 2, 2021, from https://www.bangkokpost.co.th/business/newspaper/bangkokpost

Bareggi, C. (2006). Proper words in proper places. *Lang Matters, 13,* 2-4.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243-257. https://doi.org/10.1093/llc/8.4.243

Biber, D. (2006a). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5*(2), 97-116.

Biber, D. (2006b). *University language: A corpus-based study of spoken and written registers.* Benjamin.

Biber, D. (2010). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 159–191). Oxford University Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., & Conrad, S. M. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Okesfjell (Eds.), *Out of copora: Studies in honor of Stig Johansson* (pp. 181-190). Rodopi.

Ref. code: 25636221042135YPA

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

Biber, D., Johansson, S., Leech, G., Conrad, S. M., & Finegan, E., (1999). *Longman grammar of spoken and written English*. Longman.

Byrd, P. & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5*, 31-64.

Carter, R. & McCarthy, M. J. (2001). Size isn't everything: spoken English, corpus and the classroom. *TESOL Quarterly, 35*(2), 337–40.

Chen, L. (2008). *An investigation of lexical bundles in electrical engineering introductory textbooks and ESP textbooks* [Master's thesis, Carleton University]. Library and Archives Canada.

Chun-guagng, T. (2014). An empirical research on corpus-driven lexical chunks instruction. *International Journal of English Language Teaching, 2*(2), 1-36.

Conrad, S. M. & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica, 20*, 56-71. https://doi.org/10.1515/9783484604674.56

Cowie, A.P. (1992). Multiword lexical units and communicative language teaching. In P.J.L Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 1-12). Macmillan.

Crystal, D. (2008a). Two thousand million?. *English Today, 24*(1), 3-6.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397-423.

Ref. code: 25636221042135YPA

Dastjerdi, H. V. & Rafiee, M. (2013). Corpus study of lexical bundles: Journalistic discourse in focus. *The Asian ESP Journal, 7*(3), 59-78.

Fadhila, U. (2020). Structural and functional analyses of lexical bundles in news articles written by Indonesian journalists. *Konferensi Linguistik Tahunan Atma Jaya 18*, 403-408.

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 11-23). John Benjamins.

Forbes. (n.d.). *Forbes Who-We-Are*. Forbes.com. https://www.forbes.com/connect/who-we-are/

Frankenfield, J. (2021, January 6). *Artificial Intelligence (AI).* Investopedia. https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

Güngör, F. & Uysal, H., H. (2016). A comparative analysis of lexical bundles used by native and non-native scholars. *English Language Teaching, 9*(6), 176-187.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24–44.

Hong, A. L. & Hua, T. K. (2018). Specificity in English for academic purposes (EAP): A corpus analysis of lexical bundles in academic writing. *3L: The Southeast Asian Journal of English Language Studies, 24*(2), 82-94.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–191.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2008b). As can be seen...: Lexical bundles and disciplinary variation. *English for Specific Purpose, 27*, 4-21.

Ref. code: 25636221042135YPA

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32*, 150- 169.

Jablonkai, R. (2009). "In the light of": A corpus-based analysis of lexical bundles in two EU-related registers. *WoPaLP, 3*, 1-27.

Jalali, Z. S. & Moini, M. R. (2014). Structure of lexical bundles in introduction section of medical research articles. *Procedia – Social and Behavioral Sciences, 98*, 719-726. http://doi.org/10.1016/j.sbspro.2014.03.473

Juknevičien, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. English speakers. *KALBOTYRA, 61*(3).

Kenton, W. (2021, February 2). *Forbes.* Investopedia. https://www.investopedia.com/terms/f/forbes.asp

Lewis, M. (1997). Pedagogical implications of the lexical approach. In J., Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255-270). Cambridge University Press.

Llanta, C., M. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English Academic Purposes, 14*, 84-94. http://dx.doi.org/10.1016/j.jeap.2014.01.002

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x

McCarthy, M. (1999). Turning number into thoughts. *The Language Teachers, 23*(6), 25-27.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*(3), 276-282.

Ref. code: 25636221042135YPA

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Clarendon Press.

Nekrasova, T. M. (2009). English L1 and L2 apeakers' knowledge of lexical bundles. *Language Learning Journal, 59*(3), 647-686.

O'Keeffee, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge University Press.

Pan, F., Reppen, R., & Biber, D. (2015). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes, 21*, 60-71. http://dx.doi.org/10.1016/j.jeap.2015.11.003

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Native like selection and native like fluency. In J.C. Richards & R.W. Schmidt (Eds.), *Language and Communication* (pp. 191-230). Longman.

Pang, P. (2009). A study on the use of four-word lexical bundles in argumentative essays by Chinese English majors: A Comparative Study based on WECCL and LOCNESS. *Teaching English in China, 32*, 25-45.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System, 42,* 220-331. http://dx.doi.org/10.1016/j.system.2013.12.003

Rafiee, M., Tavakoli, M., & Amirian, Z. (2011). Structural analysis of lexical bundles across two types of English news papers edited by native and non-native speakers. *MJAL, (3)*2, 136-155.

Rafiee, M., & Keihaniyan, M. (2013). A comparative analysis of lexical bundles in journalistic writing in English and Persian: A contrastive linguistic perspective.

Ref. code: 25636221042135YPA

Raupach, M. (1984). Formulae in second language speech production. In H.W. Dechert & D. Möhle (Eds.), *Second language production* (pp. 114-137). Gunter Narr Verlag.

Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contrastive rhetorical perspective. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 191-210). Cambridge University Press.

Reppen, R. (2010). Building a corpus: What are the basics?. In A. O'Keefe & M. McCarthy (Eds.), *The routledge handbook of corpus linguistics* (pp. 31-38). Routledge.

Santos, T., Atkinson, D., Erickson, M., Matsuda, P. K., & Silva, T. (2000). On the future of second language writing: A colloquium. *Journal of Second Language Writing*, *9*(1), 1-20. https://doi.org/10.1016/S1060-3743(99)00022-3

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1-22). John Benjamins.

Shin, Y. K., & Kim, Y. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System, 69*, 79-91. https://doi.org/10.1016/j.system.2017.08.002

Short, J. N., Boniche, A., Kim, Y., & Li, P. L. (2001). Cultural globalization, global English, and geography journals. *Professional Geographers, 53*(1), 1-11.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review, 64*(3), 429–458.

Steimle, J. (n.d.). *How to become a Forbes Writer*. Johnsteimle. https://joshsteimle.com/writing/become-forbes-writer.html

Ref. code: 25636221042135YPA

*The Importance of Technology in Our Daily Life – How Has Technology Changed Our Lives?*. (2019, November 21). Scientificworldinfo. Retrieved February 2, 2021, from https://www.scientificworldinfo.com/2019/11/importance-of-technology-in-our-daily-life.html

Vaughan, E. (2008). "Got a Date or Something?" An analysis of the role of humor and laughter in the workplace meetings of English language teachers. In A. Ädel and R. Reppen (Eds.) *Corpora and discourse: The challenges of different settings* (pp.95-115). John Benjamins.

*What is 5G.* (n.d.). Qualomm. Retrieved March 13, 2021, from https://www.qualcomm.com/5g/what-is-5g

Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal, 20*(1), 1-15.

Wray, A. (2006). Formulaic language. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 590–597). Elsevier.

Yoon, H. & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*, 257-283.

Yorio, C. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (pp. 55-71). Cambridge University Press.

Ref. code: 25636221042135YPA

# APPENDICES

Ref. code: 25636221042135YPA

# APPENDIX A
# INTERCODER ANALYSIS OF STRUCTURAL TYPES OF LEXICAL BUNDLES

The structural analysis of these lexical bundles from the researcher were in line with those from the intercoder. It is recommended that the minimum acceptable interrater agreement is 0.80 (80%) agreement (McHugh, 2012). The percentage of agreement was 1.00 (perfect agreement) which was considered as high reliability of the analyses.

| **Structural types** | **Coded units** | **Agreement** | **Disagreement** | **Percentage** |
|---|---|---|---|---|
| **1.Lexical bundles that incorporate *verb phrase* fragments** | **6** | **6** | **0** | **100** |
| 1a. (connector +) 1st / 2nd person pronoun + VP fragment | - | - | - | - |
| 1b. (connector +) 3rd person pronoun + VP fragment | 4 | 4 | 0 | 100 |
| 1c. Discourse marker + VP fragment | - | - | - | - |
| 1d. Verb phrase (with non-passive verb) | 2 | 2 | 0 | 100 |
| 1e. Verb phrase with passive verb | - | - | - | - |
| 1f. *yes-no* question fragments | - | - | - | - |
| 1g. WH-question fragments | - | - | - | - |

Ref. code: 25636221042135YPA

(Continued)

| Structural types | Coded units | Agreement | Disagreement | Percentage |
|---|---|---|---|---|
| **2.Lexical bundles that incorporate *dependent clause* fragments** | **2** | **2** | **0** | **100** |
| 2a. 1st / 2nd person pronoun + dependent clause fragment | - | - | - | - |
| 2b. WH-clause fragments | - | - | - | - |
| 2c. *If*-clause fragments | 1 | 1 | 0 | 100 |
| 2d. (verb/adjective +) *to*-clause fragment | - | - | - | - |
| 2e. That-clause fragments | 1 | 1 | 0 | 100 |
| **3.Lexical bundles that incorporate *noun phrase and prepositional phrase* fragments** | **7** | **7** | **0** | **100** |
| 3a. (connector +) Noun phrase with *of*-phrase fragment | 1 | 1 | 0 | 100 |
| 3b. Noun phrase with other post-modifier fragment | - | - | - | - |
| 3c. Other noun phrase expressions | - | - | - | - |
| 3d. Prepositional phrase expressions | 6 | 6 | 0 | 100 |
| 3e. Comparative expressions | - | - | - | - |
| **Total** | **15** | **15** | **0** | **100** |

Ref. code: 25636221042135YPA

# APPENDIX B
# INTERCODER ANALYSIS OF FUNCTIONAL TYPES OF LEXICAL BUNDLES

There were 4 disagreements between the functional analyses from the researcher and the intercoder. Therefore, the percentage of agreement was 0.73 (73.33%). The percent agreement of the intercoder analysis of functional types is below 0.80 (80%). The discourse functions of lexical bundles are sometimes ambiguous. Therefore, this incident may occur as the researcher and the intercoder might interpret and perceive the function in discourse of the lexical bundles differently.

| **Discourse functions categories (Biber et al., 2004)** | **Coded units** | **Agreement** | **Disagreement** | **Percentage** |
|---|---|---|---|---|
| **1. Stance bundles** | **6** | **4** | **3** | **26.67** |
| 1.1 Epistemic stance | - | - | - | - |
| 1.2 Attitudinal/modality stance | | | | |
| Desire | 0 | 0 | 1 | 0.00 |
| Obligation/directive | - | - | - | - |
| Intention/prediction | 4 | 2 | 2 | 13.33 |
| Ability | 2 | 2 | 0 | 13.33 |
| **2. Discourse organizers** | **1** | **0** | **1** | **0.00** |
| 2.1 Topic introduction | - | - | - | - |
| 2.2 Topic elaboration/clarification | 1 | 0 | 1 | 0.00 |
| **3. Referential expressions** | **6** | **5** | **2** | **33.33** |
| 3.1 Specification of attributes | | | | |
| Intangible framing | - | - | - | - |
| Tangible framing | 1 | 1 | 0 | 6.67 |
| Quantity | 0 | 0 | 1 | 0.00 |
| 3.2 Identification/focus | 1 | 0 | 1 | 0.00 |
| 3.3 Imprecision | - | - | - | - |

(Continued)

Ref. code: 25636221042135YPA

| **Discourse functions categories (Biber et al., 2004)** | **Coded units** | **Agreement** | **Disagreement** | **Percentage** |
|---|---|---|---|---|
| 3.4 Time/place/text reference | | | | |
| Place | 3 | 3 | 0 | 20.00 |
| Time | 1 | 1 | 0 | 6.67 |
| Text deixis | - | - | - | - |
| Multifunctional reference | - | - | - | - |
| **4. Special conversational functions** | **0** | **0** | **0** | **0.00** |
| 4.1 Politeness | - | - | - | - |
| 4.2 Simple inquiry | - | - | - | - |
| 4.3 Reporting | - | - | - | - |
| **Discourse functions categories (Hyland, 2008b)** | | | | |
| **1. Research-oriented** | **4** | **4** | **1** | **26.67** |
| 1.1 Location | 4 | 4 | 0 | 26.67 |
| 1.2 Procedure | - | - | - | - |
| 1.3 Quantification | 0 | 0 | 1 | 0.00 |
| 1.4 Description | - | - | - | - |
| 1.5 Topic | - | - | - | - |
| **2. Text-oriented** | **1** | **1** | **0** | **6.67** |
| 2.1 Transition signals | - | - | - | - |
| 2.2 Resultative signals | 1 | 1 | 0 | 6.67 |
| 2.3 Structuring signals | - | - | - | - |
| 2.4 Framing signals | - | - | - | - |
| **3. Participant-oriented** | **1** | **1** | **0** | **6.67** |
| 3.1 Stance | - | - | - | - |
| 3.2 Engagement | 1 | 1 | 0 | 6.67 |
| **Shared lexical bundles** | **4** | **4** | **3** | **26.67** |
| **Total** | **15** | **11** | **4** | **73.33** |

Ref. code: 25636221042135YPA

# BIOGRAPHY

| | |
|---|---|
| Name | Miss Lalitpatr Rujirawan |
| Date of Birth | March 16, 1993 |
| Educational Attainment | 2015: Bachelor of Accountancy<br>Chulalongkorn University |
| Work Position | June 2021 – Present: Vendor Key Account Management<br>Lazada Thailand, Bangkok |
| Work Experiences | December 2020 – May 2021: Assistant Business Development Manager<br>Collars Hospitality Co., Ltd., Bangkok<br>December 2016 – November 2020: Cabin Crew<br>Thai Airways International Public Company Limited, Bangkok<br>July 2015 – October 2016: Audit Assistant<br>Deloitte Touche Tohmatsu Jaiyos Co., Ltd., Bangkok |

Ref. code: 25636221042135YPA