ANALYSIS OF FOUR-WORD LEXICAL BUNDLES: AN APPLICATION TO ENHANCE PUBLIC SPEAKING SKILL BY EXTRACTING CHUNKS FROM TED-TALK

BY

MS. NUTCHADA SUWANWONG

AN INDEPENDENT STUDY PAPER SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS IN
TEACHING ENGLISH AS A FOREIGN LANGUAGE
LANGUAGE INSTITUTE
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2015

ANALYSIS OF FOUR-WORD LEXICAL BUNDLES: AN APPLICATION TO ENHANCE PUBLIC SPEAKING SKILL BY EXTRACTING CHUNKS FROM TED-TALK

BY

MS. NUTCHADA SUWANWONG

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

LANGUAGE INSTITUTE

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2015

THAMMASAT UNIVERSITY
LANGUAGE INSTITUTE

INDEPENDENT STUDY PAPER

BY

MS. NUTCHADA SUWANWONG

ANALYSIS OF FOUR-WORD LEXICAL BUNDLES: AN APPLICATION TO ENHANCE PUBLIC SPEAKING SKILL BY EXTRACTING CHUNKS FROM TED-TALK

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

on May 22, 2016

Chairman _____
(Chanika Gampper, Ph.D.)

Member and Advisor _____
(Assistant Professor Monnipha Somphong, Ph.D.)

Dean _____
(Associate Professor Pornsiri Singhapreecha, Ph.D.)

Independent Study Paper Title	Analysis of Four-Word Lexical Bundles: an Application to Enhance Public Speaking Skill by Extracting Chunks from Ted-Talk
Author	Ms. Nutchada Suwanwong
Degree	Master of Arts
Major Field/Faculty/University	Teaching English as a Foreign Language Language Institute Thammasat University
Independent Study Paper Advisor	Assistant Professor Monnipha Somphong, Ph.D.
Academic Years	2015

ABSTRACT

This study was conducted to help enhance language proficiency in speaking skill for Thai learners who are unable to speak English even they have been studying English for a long time. The real English of public speaking was examined. The paper illustrates four-word lexical chunks extracted from TED Talks. It could be said that the TED corpus, containing 500,000 words, is investigated by the AntConc application for word frequency classification. The study intends to identify the patterns of language usage in the aspects of structural and functional taxonomy. The corpus technique helps reveal the real English from real–life conversation that can be beneficially used by a teacher. The paper is meant to be the beginning of a discussion of how word frequency could be developed as the effective material, or applied for a lexical approach, to help enhance the accurate and fluent English speaking of language learners.

Keywords: Corpus, Lexical Chunk, Public Speaking

ACKNOWLEDGEMENTS

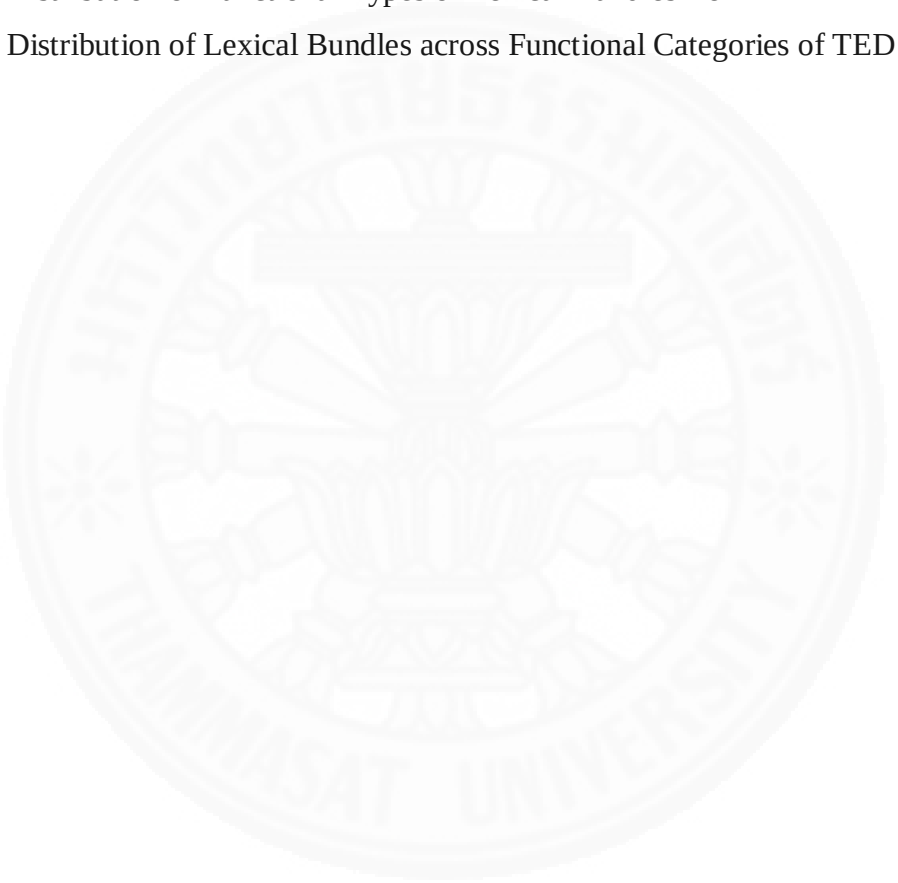
First of all, I would like to express my deepest appreciation to Asst. Prof. Dr. Monnipha Somphong, my advisor, who always guided and provided me useful advice to complete this study. A special gratitude to Asst. Prof. Dr. Supakorn Phoocharoensil, who firstly inspired me to conduct the paper based on corpus study. Finally, I must express my profound gratitude to my beloved family for providing me with great support and encouragement throughout my years of the study in TEFL, Teaching English as a Foreign Language, at Thammasat University. This achievement would not have been accomplished without them. Thank you.

Ms. Nutchada Suwanwong

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES



CHAPTER 1

INTRODUCTION

1.1 BACKGROUND TO THE STUDY

It seems inevitable that the challenge of EFL/ ESL learning is how learners are enabled to effectively automize their second language for real communication. In other words, the majority of EFL/ESL students can clearly describe the grammar rules; however, they fail to apply those rules in their own words for real life communication in speaking and writing (Randall, 2009).

According to O'Keeffe, A., McCarthy, M., Carter, R. (2007), in order to be more native-like or have native fluency, learners should expose themselves to the use of chunks for language drilling to attain the maximization of automatic second language usage. To illustrate, with over emphasis in language learning on single words, the learners tend not to be able to produce the real language fluently. As Chun-guang (2014) has indicated, the most crucial factor that enables native speakers to produce the language automatically and idiomatically is that they tend to store a large number of chunks in their brain (Becker, 1975). Referring to the study of Chung-guang (2014), lexical chunking helps enhance the native-like accuracy and fluency of speech output since the ready-made chunks stored in mind automatically integrate and generate the language production (Sinclair, 1983).

There is little doubt that in order to achieve the ultimate goal of automatic performance in speech utterance, Thai EFL learners should be provided with effective learning instruction under the innovative framework of a corpus-driven approach.

1.2 STATEMENT OF PROBLEM

Taking into EFL context for Thai learners, the studies of Boonkit (2010) and Jidathai (2015) reveal that English speaking is considered as a crucial skill of effective

communication. However, with the Thai context of EFL learning, it is highly likely that Thai undergraduate students and employed graduates are unable to speak English effectively even though they have been bombarded with English learning for almost eight to ten years. Their English proficiency is highly likely to be below the average. It appears that the cause of difficulties in English speaking could be identified as a lack of confidence in terms of anxiety about making errors, attitude, motivation, personality, or English exposure, as well as the teaching and learning English management (Boonkit, 2010 & Jidathai, 2015). Furthermore, the work of Kunnu & Sukwises (2014) shows that another main reason that English language teaching fails is the emphasis on grammar translation teaching methods. Consequently, Thai EFL learners are unable to attain conversational proficiency.

As cited in the study of Chun-guang (2014), it is quite probable that in order to produce the utterance, vocabulary plays a more important role than grammar (Wilkins, 1972). Taking into account traditional vocabulary learning, learners tend to be blasted by grammar rules and single word memorizing so they are unable to comprehend native speakers' speech conveyed by frequently used lexical chunks, and to produce these utterances effectively. Chun-guang (2014) also claims that even if learners have been exposed to the English language learning for many years, it seems that they are unable to produce the real language in speaking and writing effectively. It could be said that they tend to focus on grammar structure rather than the English word choice. As a result of this, native speakers are unlikely to understand.

It is difficult to imagine how people in Thailand can be enabled to compete with others in the highly competitive era of the incoming Asian Economic Community (AEC) where the demand for proficiency of English-speaking workers is growing. It is fairly certain that people with strong English oral communicative skills will have better job opportunities (Plangkham & Porkaew, 2012). As a result of this, it is almost inevitable that in order to enhance language proficiency in English speaking, a corpus driven approach, utilizing lexical chunks, should be implemented. As Chun-guang (2014; p.7) points out "Teaching lexical chunks can help ease learners' frustration, promote their motivation, and improve language fluency".

1.3 RESEARCH OBJECTIVE

This research was conducted to explore the discourse of TED applying a frequency-based approach advised by Biber&Conrard (2009). Thus, this paper aims to investigate the effectiveness of lexical bundles in public speaking and to explore the strategy of corpus driven learning to help enhance the English proficiency of Thai learners by implanting the lexical knowledge from public speaking of English short talk: TED Talks. The investigation of lexical bundles in the register will be analyzed into structural and discourse taxonomies perspectives as the classification of lexical bundles (Bal, 2010; Biber et al., 2004).

1.4 RESEARCH QUESTIONS

This paper is conducted to examine the effectiveness of a corpus-driven approach; the research question can be identified as below.

1. What are the most common four-word lexical bundles found in public speaking by native speakers?
2. What are the structural and functional features of lexical bundles found in the study?

1.5 DEFINITION OF TERMS

1.5.1 Lexical bundles refer to fixed or semi-fixed lexical phrases or chunks which tend to be stored and retrieved as whole units automatically at one time when producing the language.

1.5.2 An application refers to an implementation of conversational phrases for public speaking.

1.6 SIGNIFICANCE OF THE STUDY

This study was done by extracting chunks from TED Talk as authentic public speaking which aimed to explore only four-word lexical bundles from 196 TED Talks of native speakers from 2004– 2016. These results will be beneficial for English teacher by implementing these potentially four-world lexical chunks as a guideline for course development, teaching strategies and authentic teaching material to help enhance the learners' language proficiency of public speaking effectively.

1.7 ORGANIZATION OF THE STUDY

This study is divided into 5 parts as the following sequences:

1.7.1 Chapter one provides the introduction; background to the study, problem statement, research objective, research questions, definition of term, scope and significant of the study as well as organization of the study.

1.7.2 Chapter two displays the relevant significant literature review concerning the definition of corpus-based studies, lexical chunk definition and classification, lexical taxonomy and definition of short talk as public speaking.

1.7.3 Chapter three proposes the TED Talk data collection and data analysis of concordancing: AntConc and lexical bundles.

1.7.4 Chapter four presents the relevant result in the respects of Ted lexical bundles, structural and functional taxonomy analysis. Moreover, the relationship between structural and functional bundles and the discussion are provided.

1.7.5 Chapter five offers a summary of the study, implications and recommendations for further study.

CHAPTER 2

LITERATURE REVIEW

2.1 DEFINITION OF CORPUS-BASED STUDIES: DATA DRIVEN LEARNING THEORY

According to Guang (2014), it could be said that data–driven learning (DDL) based on corpus could be identified as a kind of discovery learning pattern acted out as "language detectives" implementing both deductive and inductive process (Hadley, 2002). It appears that one of the greatest advantages of DDL is to discover the authentic language.

Taking into the consideration of corpus-based study, to quote from Bal (2010, p.5), "The corpus in the Applied Linguistic could be defined as a large collection of machine-readable texts." Moreover, referring to the study of Bal (2005), the characteristics based on the corpus-based learning could be identified as three main aspects; collections of naturally occurring texts, automatic and interactive computer analysis, and language pattern interpretation derived from quantitative and functional analysis (Conrad, 1996). As cited by Yusu (2014), with the authentic data retrieved from corpora, teaching material can be effectively improved. To put it more simply, there is little doubt that corpus-based study tends to be beneficial for teachers and learners to naturally acquire language learning in terms of grammatical patterns, word usages, semantic and pragmatic features, as well as textual discourse (Flowerdew, 2009).

It is the view of Chun-guang (2014) that corpus linguistic has an impact on vocabulary instruction (Nation, 2001; Schmitt, 2000). It is quite probable that DDL methods provide us with new insights of meta-linguistic awareness. It seems that vocabulary learning involves not only learning definition, part of speech, and

pronunciation, but also semantic preference and semantic prosody (Chun-guang, 2014).

2.2 LEXICAL CHUNK

2.2.1 LEXICAL CHUNK DEFINITION AND CLASSIFICATION

As cited in the study of Chun-guang (2014), with the rising of new perspective corpus linguistic and second language acquisition study, there is little doubt that in order to produce natural speech in real life communication, the automatic language output is processed by retrieving the large amount of phrases or lexical chunks stored in memory rather than resorting to the grammar rules and individual words (McCarthy & Carter, 1997; Nattinger&DeCarrico, 1992; Altenberg, 1998 and Zimmerman, 1997). O' Keeffe at el., (2007) and Zhao (2009) also point out, the characteristic of chunks could be identified as fixed or semi-fixed lexical phrases which tend to be stored and retrieved as whole units automatically at one time when producing the language.
As cited in the study of Lewis (1993), the characteristics of lexical chunks are highly likely to be those of independent units providing pragmatic meaning within a social community. It is widely accepted that the chunks are likely to be included in a wide range of either individual words or full sentences.

In an account of lexical classification, the work of Zhao (2009) and Chun-guang (2014) shows that even though the term 'lexical chunk' seems to be defined differently in various classifications, the most typical and classical one is identified by Lewis (1997b) and Nattinger & DeCarrico's (1992) as below.

In terms of Lewis's (1997b) classification, the lexical chunk can be categorized into four aspects; words and poly-words, collocations, institutionalized utterances and sentence frames and heads.

1. Words and poly-words: The words are strictly combined and cannot be replaced by the others, e.g., *out of the question.*

2. Collocation: the relationship between words that co-occur such as *vice president, deputy manager*. It could be said that this co-occurrence is relatively flexible compared to poly-words.
3. Institutionalized utterance: It is highly likely to be used in oral speech, which performs the pragmatic function. For example, *I'm afraid of…, I'm sorry but…*
4. Sentence frameworks: The difference that can define this chunk from the type three is that institutionalized utterance tends to refer to spoken language while sentence frames and heads tend to be written as organized text. The samples could be identified as *on the one hand, on the other hand.*

In respect of Nattinger's and DeCarrico's (1992) classification, lexical chunk can be defined as four types; poly-words, institutionalized expressions, phrasal constraints, and sentence builders.

1. Poly-words: The chunks refer to short phrases that cannot be replaced by other words.
2. Institutionalized expression: These unchangeable and unbroken constituents are highly likely to be memorized as whole chunks with meaningful units. For instance:
   I don't know if..., it's possible to...
3. Phrasal constraints: The different lengths from short to medium phrases. For example, noun phrases, verb phrases, and adjective phrases, e.g., *at the end of, in order to.*
4. Sentence builders: These phrasal expressions outline the structures when making explicit the different ideas as whole sentences. For example, *There is no doubt that..., what I mainly wanted to talk to you were about….*

2.3 LEXICAL TAXONOMY

Looking into the studies most relevant to this conducted research, the prominent perspective of lexical analyses could be identified into two main aspects; structural form and discourse function (Conrad & Biber, 2005 and Bal, 2010). The work of Bal (2010) reveals that a large number of multi word expressions seem to focus on the significant of structural and functional taxonomies (Altenberg, 1998; Biber,

Johansson, Leech, Conrad, & Finegan, 1999; Nattinger & Decarrico, 1992; Pawley & Syder, 1983). In the light of Jablonkai's (2009) work, the significance of a frequency based-approach in each register is highly likely to provide us the individual bundles of lexical chunks. There is little doubt that the typical and unique communicative purpose of the register is differentiation. The present study tends to investigate the register of public speaking known as TED.

2.3.1 STRUCTURAL TYPES OF LEXICAL BUNDLES

As cited in the study of Biber et al., (2004), it is highly unlikely that most lexical bundles that occur in face to face conversation are not idiomatic in meaning. Moreover, it is very doubtful if most lexical bundles represent a complete structural unit. In many cases, most research tends to ignore lexical blocks that cut across the grammatical phrases and structures by ignoring the discourse function, which is considered as a prominent communicative part of speaking (Biber et al., 2004).

Moving into a word frequency perspective of lexical chunks, Conrad and Biber (2005) and Bal (2010) have drawn attention to the fact that the structural form of lexical bundles based on the Longman Grammar of Spoken and Written English were categorized into 12 structural types. However, the study of Bal (2010) has applied the model and placed them into two main categories; phrasal and casual. The phrasal bundles could be subcategorized into Noun-Phrase (NP) based such as *the end of the* or *the way in which*, Preposition Phrase (PP) based such as *at the same time* or *on the other hand*, and Verb Phrase (VP) based such as *it is possible to* or *is one of the*. For clausal lexical bundles, one broad group of a verb/adjective followed by a to – clause fragment such as *to be able to* and adverbial clause such as *if there is a* is prominent. Lastly, a verb phrase followed by that - clause fragment such as *should be noted that* is identified.

The work of Biber et al., (2004) indicates that although most lexical bundles tend to bridge two structure of two clauses, they have strong grammatical correlations that could be identified as three main structural types, which are lexical bundles that incorporate verb phrases fragments, lexical bundles that incorporate dependent clause

fragments, and lexical bundles that incorporate noun phrases and preposition phrase fragments. To illustrate, the structural type of bundles can be viewed in Table 1.

Table 1: Structural type of lexical bundles

Structural types	**Sub-types**	**Sample bundles**
1. Lexical bundles that incorporate verb phrase fragments	1.a 1st/2nd person pronoun + VP fragment	*I'm not going to*
	1.b 3rd person pronoun + VP fragment	*and this is a*
	1.c discourse marker + VP fragment	*I mean I don't*
	1.d Verb phrase (with non-passive verb)	*have a lot of*
	1.e Verb phrase (with passive verb)	*is based on*
	1.f yes-no question fragments	*are you going to*
	1.g Wh-question fragments	*what do you think*
2.Lexical bundles that incorporate dependent clause fragments	2.a 1st/2nd person pronoun + dependent clause fragment	*I want you to*
	2.b WH-clause fragment	*when we get to*
	2.c If-clause fragment	*if we look at*
	2.d to-clause fragment	*to be able to*
	2.e that-clause fragment	*that this is a*
3. Lexical bundles that incorporate noun phrase / preposition fragments	3.a Noun phrase with of -phrase fragment	*one of the things*
	3.b Noun phrase with other post-modifier fragment	*the way in which*
	3.c Other noun phrase expressions	*a little bit more*
	3.d Prepositional phrase expressions	*at the end of*
	3.e Comparative expressions	*as well as*

Note. Reprinted from "IN THE LIGHT OF": A Corpus–Based Analysis of Lexical Bundles in Two EU-Related Registers (p.5-6) by Jablonkai, 2009.Corvinus University of Budapest :WopaLP.

According to the structural classification, there are three main structural types identified by Biber et al.,(2004). Firstly, lexical bundles that incorporate verb phrase fragments are divided into seven sub- categories; 1.a and 1.b) beginning with a subject pronoun followed by a verb phrase (*e.g. I'm not going to*), 1.c) beginning with a discourse marker followed by a verb phrase (*e.g. and this is a*), 1.d and 1.e) beginning with a verb phrase (*e.g. is based on*), as well as 1.f and 1.g) beginning with a wh-question fragment (*e.g. what do you think*).The second structural types which incorporate dependent clause fragments are sub-classified as 2.a) beginning with a pronoun followed by dependent clause fragment (*e.g. I want you to*), 2.b) introducing

with a wh-clause (*e.g. when we get to*) and 2 c) if clause (*e.g. if we look at*), as well as 2.d and 2.e) beginning with a complementizer or subordinate (*e.g. to be able to*). Finally, the third type of structural bundles tend to have casual components: 3.a -3.c) consisting of noun phrases (*e.g. one of the things,* the way in which), 3.d) consisting of prepositional phrases (*e.g. a little bit more*), and 3.e) incorporating with a comparative expression (*e.g. as well as the*).

2.3.2 FUNCTIONAL TAXONOMY OF LEXICAL BUNDLES

With regards to functional taxonomy, the work of Biber et al., (2004), Conrad & Biber (2005) and Bal (2010) shows that the functional lexical bundles could be defined into three main categories; stance expressions, discourse organizers, and referential expressions as below in Table 2.

Table 2: Discourse functions of lexical bundles

Categories	**Sub-categories**	**Sample bundles**
I. Stance bundles	A. Epistemic stance	*the fact that the, and I think that*
	B. Attitude/Modality/Stance	*what do you want*
	B1) Desire	*it is important to*
	B2) Obligation/ Directive	*it's going to be*
	B3) Intention / Prediction	*it is important to*
	B4) Ability	*it is possible to*
II. Discourse Organisers	A.Topic Introduction	*in this chapter we*
	B. Topic elaboration/ Clarification	*on the other hand*
III. Referential Bundles	A. Identification Focus	*is one of the*
	B. Imprecision	*or something like that*
	C. Specification of Attributes	
	C1) Quantity specification	*a lot of people*
	C2) Tangible framing	*in the form of*
	C3) Intangible framing	*on the basis of*
	D. Time/ Place / Text Reference	
	D1) Place reference	*in the United States*
	D2) Time reference	*at the same time*
	D3) Text-deixis	*as shown in figure*
	D2) Multi-functional reference	*in the middle of*
IV. Special Referential		

Note. Reprinted from "IN THE LIGHT OF": A Corpus –Based Analysis of Lexical Bundles in Two EU-Related Registers (p.5-6) by Jablonkai, 2009.Corvinus University of Budapest :WopaLP.

Firstly, stance bundles could be defined as groups of words identifying attitude, judgment, and perspective of speaker, or writer's certainty or uncertainty. The bundles could be classified as both personal and impersonal. As cited in the study of Biber et al., (2004) stance bundles could be divided into two main categories; epistemic stance and attitudinal/modality stance. Epistemic stance provides the information framework interpreting certain, uncertain, or possible *(e.g. I don't know if, I don't think so).* Attitudinal/ Modality stance bundle is highly likely to deliver the speakers' attitudes towards the actions or events *(e.g. I want you to, I'm not going to).* Four main subcategories could be defined as desire bundles, obligation/directive bundles, intention/prediction bundles, and ability bundles.

Next, "discourse organizer" tends to help compose its structure as a topic introduction (*e.g. what do you think, do you know what*). Two main functions could be identified as topic introduction and topic clarification. As cited in the study of Biber et al., (2004), topic focus is highly likely to provide a signal when a topic is introduced in which the bundles tend to occur with first and second pronouns. For topic elaboration, the bundles such as *you know, I mean, on the other hand,* are used for more clarification.

Furthermore, "referential expressions" are highly likely to relate to number, amount, size, and number *(e.g. at the same time, at the end of the).* Four main sub-categories could be identified as identification, imprecision indicators, specification of attributes as well as time and place reference. (Biber et al., 2004). It should be noted that this study tends to investigate the functional taxonomy of lexical bundles that play as a crucial discourse-signaling role in public speaking (Conrad &Biber, 2005; Bal, 2010). However, the study of Conrad & Biber (2005) and Bal (2010) revealed a new section of functional type of special conversational functions. There is little doubt that this function tends to cover all the three traditional categories that occurred only in the

conversational register: introducing politeness *(e.g. thank you very much),* enquiry *(e.g. what are you doing),* and reporting clause *(e.g. I said to him).* To put it more simply, the discourse functional analysis in this study can be examined in the four main categories shown in Table 3.

Table 3: Discourse functions of lexical bundles in TED corpus

Categories	**Sub-categories**
I. Stance bundles	A. Epistemic stance
	B. Attitude/Modality/Stance
	B1) Desire
	B2) Obligation/ Directive
	B3) Intention / Prediction
	B4) Ability
II. Discourse Organisers	A.Topic Introduction
	B. Topic elaboration/ Clarification
III. Referential Bundles	A. Identification Focus
	B. Imprecision
	C. Specification of Attributes
	D. Time/ Place / Text Reference
	D1) Place reference
	D2) Time reference
	D3) Text-deixis
	D2) Multi-functional reference
IV. Special Referential	

2.4 DEFINITION OF PUBLIC SPEAKING AND SHORT TALK

To quote from Hamilton and Parker (1996, p. 4): "Public speaking can be identified as the process of people sharing thoughts, ideas, and feelings with each other in a commonly understandable way." Furthermore, the work of Plangkham and Porkaew (2012) indicates that public speaking is highly likely to occur with preparation in order to perform a speech in front of an audience. It seems that the aim of public

speaking is to inform, persuade, and entertain an audience without being interrupted (Sellnow, 2005; Jaffe, 2007). It is almost certain that oral skill problems in public events tend to be diminished when oral skill is learnt and practiced (McKerrow et al., 2003).

Regarding language productive performance, it is fairly certain that there is a correlation between speaking and writing from various aspects; lexical feature, sentence length and sentence structure. To illustrate further, the lexical pattern of a president's speech or a presentation from a scientist group would probably be similar to a written work (Wang, 2012). With the definition of short talk, the work of Wang (2012) reveals that a successful short talk - a less than twenty minute presentation - tends to require a more sustained level of clarity throughout the whole presentation parts; beginning, middle, and ending sessions. For the implementation of this study, the transcripts of seventy-three short (less than twenty minutes duration) public speaking items,, in TED Talks, were selected and used for TED corpus building.

2.5 PREVIOUS RESEARCH

Wang (2012) has conducted research for an exploration of vocabulary knowledge in English short talks. It is highly likely that such research can discover the word patterns, features, and usage that are highly likely to be frequently encountered by language users for their real life communication. As a result of this, a particular corpus TED was deployed through less than twenty minutes English short talks from the TED Talks website. Moreover, the data of talks longer than twenty minutes, from the existing corpus BASE, (British Academic Spoken English) were also collected. With the application of corpus tools, Antconc (Anhony, 2003), RANGE (Nation & Heartkey, 2003), and KfNgram (Fletcher, 2007), the findings show the significant, typical and frequent individual and lexical bundles of word usages throughout the beginning, middle, and ending part of English short talks. These should be beneficial for specific teaching purposes that enable teachers to build their own corpus to meet the specific need of learners (Wang, 2012).

Going beyond studies that focus on spoken discourse, Bal (2010) published a study of four-word analysis of lexical chunks in published research written by Turkish scholars who were non-native English speakers. The aim of the study was to investigate the structures and discourse functions of four-word lexical bundles by implementing the well-known software AntConc. The TSRAC bundles were elaborately analyzed and classified into the given structural and functional taxonomies. The result has shown the beneficial frequency usage of lexical bundles in academic writing registers that should be highlighted and implemented through the English academic writing purpose. Furthermore, referring to the study of Bal (2010), Cortes (2008) has compared the lexical bundles of English and Spanish historical journals. The aim of the study was to explore and identify the structural and functional discourse of its register. Besides, Kim (2009) has investigated Korean lexical bundles in conversation and academic texts to discover the crucial Korean expressions within the discourse function frame work (Bal, 2010).

From these studies of lexical bundles, it is evident that the studies of lexical bundles have been increasingly investigated in different registers: spoken vs written register, academic vs non-academic register, and English vs other language. These studies are highly likely to be explored to serve different pedagogical purposes (Bal, 2010). So far, the lexical bundles used in public speaking have been rarely investigated.

CHAPTER 3

METHODOLOGY

3.1 DATA COLLECTION

In order to build the TED corpus for public speaking purpose, the current study collected data from English public speaking considered as a short talks form - TED (www.ted.com).

3.1.1 TED TALK

TED is a non-profit organization started in 1984 and devoted to spreading ideas covering almost all topics and global issues such as technology, design, entertainment, business etc., in more than one hundred languages. It is widely known that TED, a global community, is highly likely to welcome all people from today's world of discipline and cultural differences. They all aim to seek a deeper understanding of community and help share ideas in a global world (https://www.ted.com/about/our-organization).

In the aspect of corpus size, the work of Bal (2010) indicates that in order to represent authentic word frequency in the study, a corpus size must be large enough (Biber, 2006). Bal (2010) points out that a corpus size of a grammatical structures, noun or verb, could be smaller since these corpuses are highly likely to occur. Nevertheless, for the target of other studies, a larger corpus is required (Bal, 2010).

Considering the application of this study, in order to uncover the features and pattern of English vocabulary usage in English public speaking, a collected TED corpus of one hundred and ninety six transcripts of native speakers from TED's website were deployed. Applying a corpus tool known as AntConc (Antony, 2014), the aim was to unveil the functional pattern of lexical bundles of English usage in English public

speaking. Subsequently, suggestions for teaching and learning of how to develop English public speaking could be provided.

The themes of TED for data collection in this study were assigned by the researcher. One hundred and ninety six TED Talks from the period of 2004 -2016 were collected from different 23 themes: activism, adventure, art, beauty, body language, business, creativity, communication, culture design, entertainment, family, food, global issue, health, history, love, music, psychology, science, and technology, each with the average length of fifteen minutes (Table 4).

Table 4: Theme and Description of TED Corpus

	Theme	Number of Talk	Minutes	Average Lengths	Total Token
1	Activism	4	61.81	15.45	
2	Adventure	2	31.96	15.98	
3	Arts	8	104.01	13.00	
4	Beauty	5	56.78	11.36	
5	Body Language	4	54.37	13.59	
6	Business	18	272.04	15.11	
7	communication	13	200.16	15.40	
8	creativity	5	89.61	17.92	
9	Culture	17	269.36	15.84	
10	Design	7	100	14.29	
11	Education	4	63.54	15.89	
12	Entertainment	31	458.16	14.78	
13	Environment	3	48.11	16.04	
14	Family	2	36.11	18.06	
15	Food	5	72.49	14.50	
16	Global Issue	10	158.6	15.86	
17	Health	18	272.11	15.12	
18	History	2	36.42	18.21	
19	Love	2	29.63	14.82	
20	Music	2	31.68	15.84	
21	Psychology	7	121.69	17.38	
22	Science	6	98.1	16.35	
23	Technology	21	289.83	13.80	
Total		**196**			**506816**

3.2 DATA ANALYSIS

In order to underpin data analysis, the application of an AntConc identifying concordance line would be first identified, followed by the lexical bundle implementation.

3.2.1 Concordancing: AntConc

As cited in the study of Yusu (2014), the unique features of concordance lines help enhance learners' awareness of language usage. As a result of this, autonomy of lexical output can be promoted Yusu (2014). Taking into account the aim of this study, a free useful text analysis known as AntConc (Anthony 2014) was used. It is widely known that one feature of the application provides the key word frequency generator that is crucial for word classification. Hence, it ensures that the most frequent word occurrence would be ranked on the top of a word list followed by the less frequent words (Bal, 2010; Wang, 2012).

It is crucial to state that this present study aims to find the most four common lexical bundles in TED Talks. As Bal (2010) states, "Four word lexical bundles are more common and present a wider range of structures and functions (p.17)". Also, the criteria of frequency and range cut–off points that tends to be different from study to study are identified based on the work of Biber & Conrad (2005) and Bal (2010).
In article by Biber and Conrad (2005), and Bal (2010), the criteria for frequency of four-word lexical bundles should be significant appearance in the register at least ten times per million words and occurrence in at least five different texts (Biber et al., 1999).

3.2.2 LEXICAL BUNDLES

In order to investigate lexical bundles, AntConcis is considered as an efficient tool for word clusters and N-grams analysis. As Wang (2012) points out, in order to investigate the lexical bundle, N-gram and clusters function would be deployed to

unveil the frequent patterns that are recognized as common phrases (Biber et al., 2004; Hoey, 2005; Carter & McCarthy, 2006). There is little doubt that an effective free software program helps facilitate and generate the most frequent word combination of lexical bundles used in English public speaking.

Referring to the study of Biber et al., (2004), "Lexical bundles are simply identified as the most frequent recurring lexical sequence in a register" (p.376). It seems that these high frequency patterns investigated by researchers are a prominent reflection of words stored and used as a prefabricated chunk.

In order to utilize the AntConc for data analysis, plain text is required. As a result of this, all the one hundred and ninety six articles from TED Talk were saved and uploaded via AntConc. Next, for retrieving the lexical bundles from those files, the "Cluster/N-Grams" feature was executed by setting the frequency counts as 5-grams. This is illustrated as shown in Figure 1.

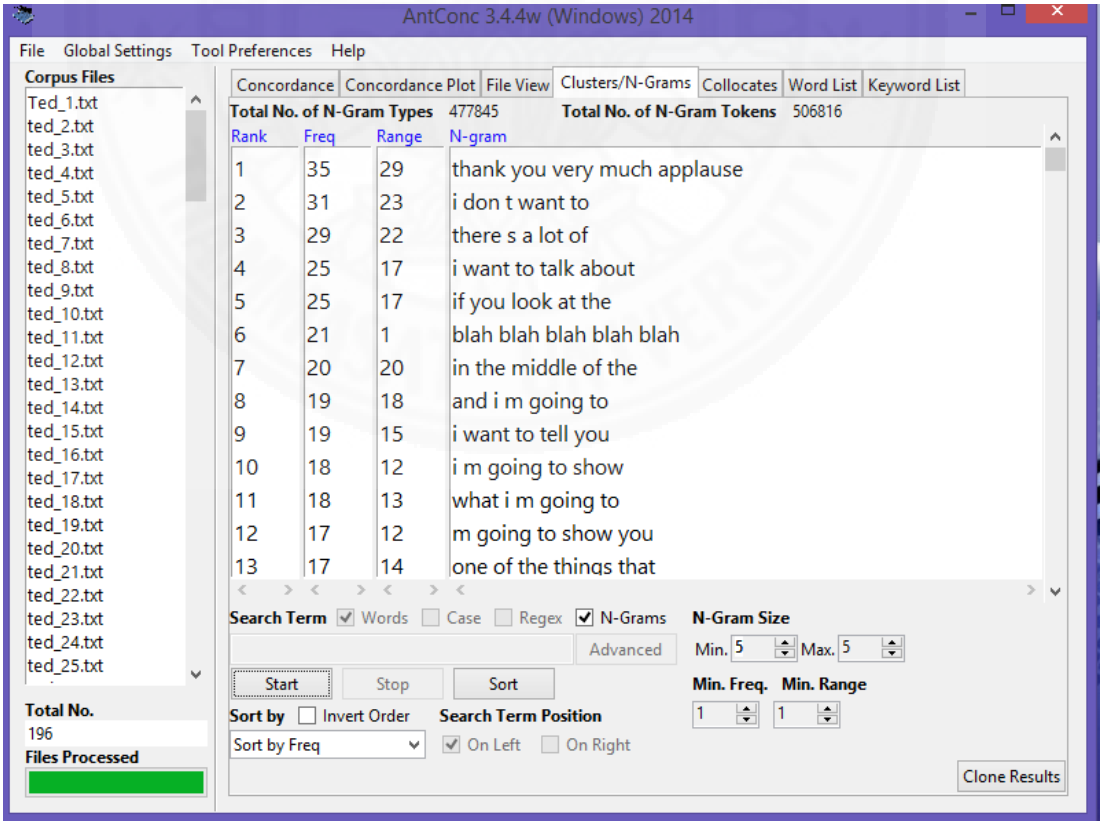


Figure 1. Data analysis implementing Clusters/ N-Grams feature of AntConc

The N-gram Size of Min and Max is set as five for four-word lexical chunks extraction due to the fact that when setting the N-gram size as four, the results seem to be deviant. To be more precise when running the results of four-size N-grams, the results such as "We're going to, I don't know, I'd like to", considered as three- word lexical chunks, were included. However, significant previous research has revealed that in order to clearly identify lexical bundles, the orthographic word units should be emphasized as one single word; for examples, don't, we're, I'd, and it's (Biber et., al, 2004).

Based on the study of Biber et. al., (2004) and Jablonkai (2009), the four words lexical bundles should be considered when they occur at least 40 times per million. Nevertheless, due to the fact that the present study represented the small corpora of TED, the following requirement is defined: only recurring four-word lexical items occurring at least five times in five different texts were included in the analysis of TED corpus, in order to overcome idiosyncratic language (Jablonkai, 2009).

After running AntConc with five N-Gram size, a list of four-word chunks was retrieved. The total number of N-Gram Types was 506816. However, the time limitation for this research study was a factor. Therefore, in order to meet the previously established cut-off points of the frequency approach of 506816 TED corpus, the first 164N-Gram types considered as a lexical bundle (occurred at least five times in at least five texts) were manually checked. There is no doubt that the insignificant duplicated expressions that appeared in less than five texts were eliminated. After that, each bundle of four-word lexical chunks were grouped referring to their structures and functionalities. The most frequent bundles found in TED were " *thank you very much, I don't want to, there's a lot of, I want to talk, if you look at, in the middle of, and I'm going to, I want to tell, I'm going to show, and what I'm going to*". However, in this study, all the bundles in TED can be illustrated as shown in Table 5.

Table 5: All lexical chunks from TED Talks

Rank	Freq	No. ofTED Talks	Cluster
1	35	29	thank you very much
2	31	23	i don t want to
3	29	22	there s a lot of
4	25	17	i want to talk
5	25	17	if you look at
6	20	20	in the middle of
7	19	18	and i m going to
	19	15	i want to tell
9	18	12	i m going to show
	18	13	what i m going to
11	17	12	I'm going to show
	17	14	one of the things
13	16	14	i m not going to
14	15	14	at the end of
15	14	11	and i think that s
	14	12	i m going to tell
	14	9	in the united states
	14	12	I'm going to tell
	14	12	so i m going to
	14	13	you don t have to
21	13	12	i don t know if
	13	9	i m going to talk
	13	11	the rest of the
24	12	11	i m going to do
	12	8	i m going to try
	12	10	it s going to be
	12	11	that i m going to
	12	10	we re going to have
	12	8	what s going to happen
	12	7	you re going to see
	12	9	you re not going to
32	11	9	and it s not just
	11	9	i don t know what
	11	9	if you think about
	11	10	it s not going to
	11	10	now i m going to
	11	6	we don t want to
	11	10	we re going to be
39	10	8	all of a sudden
	10	10	if you look at
	10	10	and that s what i
	10	9	to be able to

	10	6	ican t tell you
	10	9	i have to tell
	10	9	s not going to be
	10	10	that s what i m
	10	6	that s what we re
	10	10	this is one of
	10	9	what we re going to
50	9	8	and i think it s
	9	7	and that s what we
	9	8	how many of you
	9	7	i don t have a
	9	8	is going to be a
	9	9	it doesn t have to
	9	9	it s not just the
	9	7	that s why we re
	9	6	to figure out how
	9	9	we don t have to
	9	8	what s going on in
	9	9	when you think about
62	8	7	at the end of
	8	8	and i don t know
	8	8	doesn t have to be
	8	8	i don t know how
	8	8	is one of the most
	8	7	going to have to
	8	7	so we re going to
	8	7	this is going to
	8	6	what i d like to
	8	8	you don t need to
72	7	7	a few years ago
	7	6	and i d like to
	7	6	and we re going to
	7	7	and you re going to
	7	5	as you can see
	7	6	when it comes to
	7	7	don t get me wrong
	7	7	don t have to be
	7	5	don t know how to
	7	7	don t want to be
	7	5	don t want to get
	7	7	for the first time
	7	7	i don t have to
	7	6	i don t think this
	7	6	ill tell you what

	7	7	i want to show
	7	7	i want to talk to
	7	5	if we re going to
	7	5	let s look at the
	7	5	let s take a look
	7	7	the other side of
	7	5	there s going to be
	7	7	when you look at
	7	5	you re going to do
96	6	6	at the same time
	6	5	and i don t want
	6	6	i think this is
	6	5	i thought to myself
	6	6	it turns out that
	6	6	and that s why we
	6	6	this is a little
	6	5	are we going to
	6	6	but here s the thing
	6	6	but i think it s
	6	5	can t tell you how
	6	6	nothing to do with
	6	5	i don t want you
	6	6	i m going to ask
	6	5	i m going to be
	6	5	i m going to have
	6	6	i m just going to
	6	6	i think it s a
	6	6	if you don t know
	6	6	it turned out to be
	6	6	now we re going to
	6	6	one of the things
	6	6	so that s what we
	6	6	end of the day
	6	6	the good news is that
	6	6	this is the kind of
	6	6	to be one of the
	6	5	to come up with
	6	5	we don t know what
	6	6	we re going to see
	6	5	you can see it s
	6	6	you don t want to
	6	6	you re going to go
	6	5	you re never going to
131	5	5	and i don t think

	5	5	and i have to tell
	5	5	and it s not that
	5	5	and what that means is
	5	5	but i m going to
	5	5	don t know about you
	5	5	for a long time
	5	5	for those of you
	5	5	i d like to do
	5	5	i d like to show
	5	5	ididn t want to
	5	5	i don t know about
	5	5	i m going to give
	5	5	i m here to tell
	5	5	i said i don t
	5	5	i think we have to
	5	5	i think we re going
	5	5	i want to share with
	5	5	is that there s a
	5	5	it s kind of like
	5	5	it turns out it s
	5	5	it was the first time
	5	5	so that s what i
	5	5	so what i m going
	5	5	spent a lot of time
	5	5	thank you so much
	5	5	there s a couple of
	5	5	want to share with
	5	5	want to talk about
	5	5	we re going to do
	5	5	we re going to get
	5	5	we ve been able to
	5	5	what s happening in the
	5	5	what that means is
	5	5	what we find is

CHAPTER 4

RESULTS AND DISCUSSION

This chapter reveals the most common four word lexical bundles found in the public speaking of TED by native speakers, and exposes the structural and functional features in the spoken register. The lexical bundles identified in the TED corpus are introduced. Furthermore, the result of the quantitative and qualitative analysis will be clarified followed by a discussion.

4.1 TED LEXICAL BUNDLES

This chapter introduces the lexical bundles identified in TED. Moreover, the results of quantitative and qualitative data will be analyzed as a discussion for these results. A total of one hundred and thirty eight items were identified in the TED. The top ten most frequent ranks of bundles found in TED were: "*thank you very much, I don't want to, there's a lot of, I want to talk, if you look at, in the middle of, and I'm going to, I want to tell, I'm going to show, and I'm going to*". However, since the purpose of the study is to discover the structural and functional features of lexical bundles in TED, the detailed analysis of these two categories will be provided through the study of all one hundred and thirty eight bundles.

4.2 STRUCTURAL TAXONOMY ANALYSIS

As shown in Figure 1, structural taxonomy analysis of TED, the largest part of 95 lexical bundles is comprised of lexical bundles that incorporate verb phrase fragments (69 percent) followed by 26 lexical bundles (17 percent) that incorporate dependent clause fragments respectively. Finally, 17 bundles are formed by noun and prepositional phrase fragments (12 percent). The result of the analysis in this study is summarized in Figure 2.

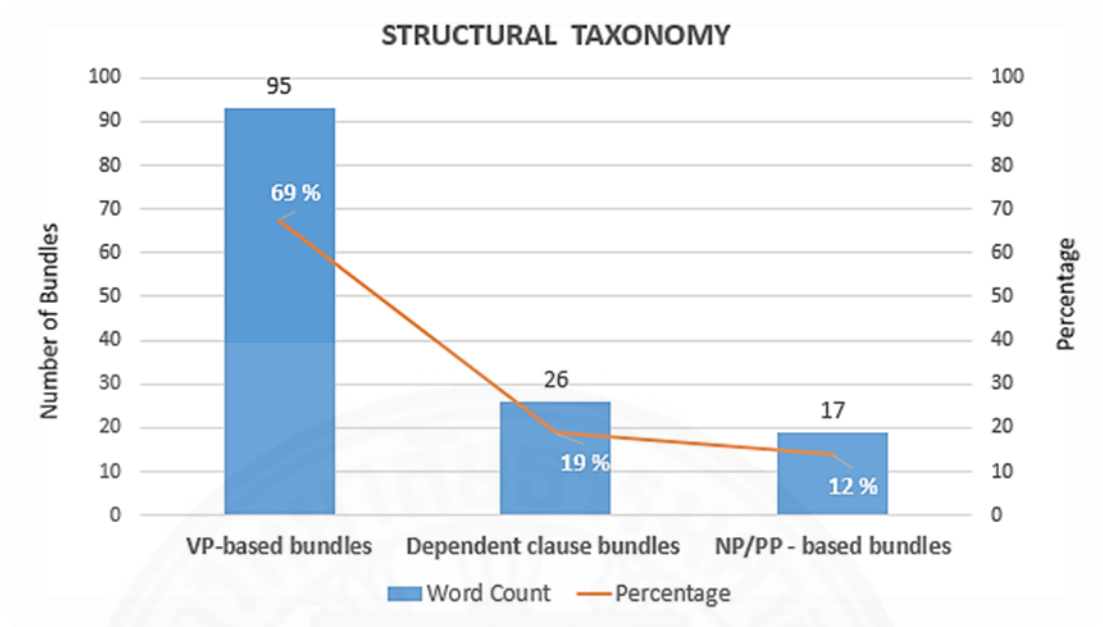


Figure 2. Structural taxonomy analysis of TED

In the study, the lexical bundles in TED are deliberately exploratory to group the bundles according to their structural characteristics as defined by Biberet. al., (2004). Undoubtedly, there are three broad types of categories: lexical bundles that incorporate verb phrase fragments; lexical bundles that incorporate dependent clause fragments, and lexical bundles that incorporate verb phrase fragments. To illustrateI this more simply, all of the structural bundles are elaborately classified in Table 6.

Table 6: Structural taxonomy of four-word lexical chunks from TED

I	**Lexical bundles that incorporate verb phrase fragments**				
1.a	1st/2nd person pronoun + VP fragment	I don't want to	I want to show	you don't have to	you don't want to
		I want to talk	I didn't want to	we don't have to	I have to tell
		I want to tell	I don't have a	I don't have to	I can't tell you
		we don't want to	I don't want you	we've been able to	but I want to
		and I'm going to	I'm going to tell	we're going to have	I'm going to ask
		what I'm going to	so I'm going to	you're going to see	I'm going to be
		I'm not going to	I'm going to talk	you're not going to	I'm going to have

		now I'm going to	I'm going to do	we're going to be	and I'd like to
		but I'm going to	and we're going to	so we're going to	I'd like to do
		that I'm going to	and you're going to	you're going to do	I'd like to show
		I'm going to show	I'm just going to	now we're going to	I'd like to talk
		we're going to see	we're going to do	I'm going to get	I'm here to tell
		you're going to go	we're going to get	I'm going to give	you can see it's
		you're never going to	and I don't think	as you can see	are we going to
		and I think that's	and I think it's	we don't know what	
1.b	3rd person pronoun + VP fragment	it doesn't have to	this is going to	so that's what I	and it's not just
		it's not just the	there's going to be	but here's the thing	this is one of
		it's going to be	and this's not that	it turns out it's	there's a couple of
		it's not going to	going to be able	it turns out that	there's a lot of
		is going to be	it's kind of like	it turned out to	this is a little
1.c	discourse marker + VP fragment	I think this is	I said I don't	I think we have to	but I think it's
1.d	Verb phrase (with non-passive verb)	don't get me wrong	don't have to be	want to share with	don't want to be
		can't tell you how	don't know how to	want to talk about	don't want to get
		let's look at the	let's take a look	to come up with	to be one of
1.e	Verb phrase (with passive verb)				
1.f	yes-no question fragments				
1.g	Wh-question fragments				
II	**Lexical bundles that incorporate dependent clause fragments**				
2.a	1st/2nd person pronoun + dependent clause fragment	I don't know if	I don't know how	I'll tell you what	
		I don't know what	I thought to myself	I don't know about	

2.b	WH-clause fragment	what I'd like to	what that means is	how many of you	
		what we're going to	what's going on in		
		when you think about	when it comes to		
		what's happening in the	what that means is		
		what's going to happen	what we find is		
2.c	If-clause fragment	if you look at	if we're going to		
		if you think about	if you don't know		
2.d	to-clause fragment	to figure how to	to figure out what		
2.e	that-clause fragment	is that there's a			
III	**Lexical bundles that incorporate noun phrase / preposition fragments**				
3.a	Noun phrase with of -phrase fragment	one of the things	one of the most	all of the sudden	
		the other side of	the rest of the	end of the day	
3.b	Noun phrase with other post-modifier fragment				
3.c	Other noun phrase expressions	nothing to do with	thank you very much		
		a few years ago	thank you so much		
3.d	Prepositional phrase expressions	for those of you	for the first time	at the same time	
		to come up with	in the same way		
		in the middle of	for a long time		

		at the end of	in the United States		
		to be one of	at the end of		
3.e	Comparative expression				

As cited in the study of Biberet. al., (2004), most lexical bundles in a conversational register are likely to be uncompleted structural units. It is crucial to state that the boundary of conversational structural units tends to begin with a clause or phrase in which the word ending is considered as the first hint of the second structural unit. To illustrate, the discourse marker followed by VP fragment of *I mean you know, you know it was,* are well represented as samples (Biber et. al., 2004, p. 381). Besides, a large number of bundles tend to be compiled in two clauses; such as *I want to know, well that's what I.* On the contrary, bundles in an academic register tend to bridge two phrases such as *in the case of, the base of the* (Biber et. al., 2004, p. 377). In parallel with what Biber et al., (2004) found, Conrard & Biber (2005), and Bal (2010) said the lexical bundles found in TED are highly likely to be ungrammatical complete units.

4.3 FUNCTIONAL ANALYSIS

As explained in Chapter 2, the functional taxonomy in this study can be distributed into four categories: stance bundle, discourse organizer, referential expression and special conversational functions. To be more analytical, analysis of the bundles was subcategorized for each section (Biber et. al., 2004). The functional classification of all the four–word lexical bundles identified in the TED will be presented in Figure 3.

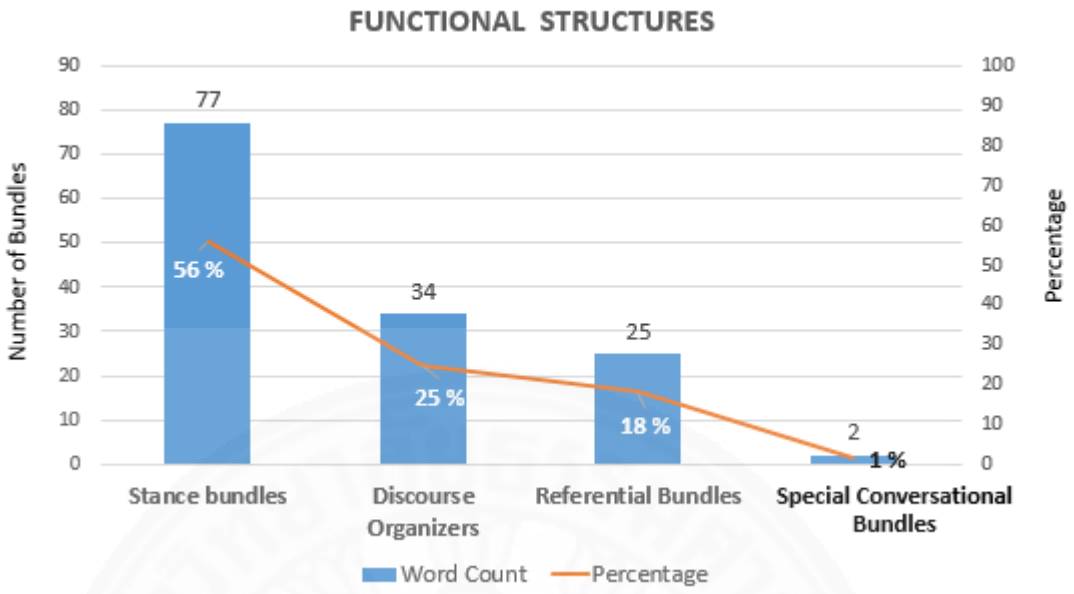


Figure 3. Distribution of functional types of lexical bundles from TED

The findings show that spoken language in public speaking from TED talks is dominated by 77 stance expressions (56 percent) followed by 34 discourse organizers (25 percent), 25 referential expressions (18 percent) and 2 special conversational functions (1 percent), respectively. As cited in the study of Jukneviciene (2009), the characteristics of spoken language could be identified as stance expressions and discourse organizing, while the characteristics of written academic could be identified as referential expression (Biber et al., 1999, 2004; Biber 2006; O'Keeffe et al., 2007). The functional bundles are categorized as in Table 7.

Table 7: Functional discourse of four-word lexical chunks from TED

I	**Stance Expression**				
	A	Epistemic stance			
		Personal	and I think that's	but I think it's	I think we have to
			and I think it's	and I don't think	I don't know how
			I don't know if	I think this is	I don't have a
			I don't know what	we don't know what	I don't know about
		Impersonal			
	B	Attitude/Modality/Stance			

		B1. Desire			
		Personal	I don't want to	want to share with	
			I want to talk	but I want to	
			I want to tell	don't want to be	
			we don't want to	don't want to get	
			I want to show	we've been able to	
			I didn't want to	want to talk about	
		Impersonal			
		B2. Obligation/ Directive			
		Personal	you don't have to	I have to tell	can't tell you how
			we don't have to	don't get me wrong	don't have to be
			I don't have to	I said I don't	don't know how to
			I don't want you	I can't tell you	you don't want to
		Impersonal	it doesn't have to	it's not just the	
		B3. Intention/ Prediction			
		Personal	and I'm going to	I'm going to tell	we're going to have
			what I'm going to	so I'm going to	you're going to see
			I'm not going to	I'm going to talk	you're not going to
			now I'm going to	I'm going to do	we're going to be
			but I'm going to	and we're going to	so we're going to
			that I'm going to	and you're going to	you're going to do
			I'm going to show	I'm just going to	we're going to do
			we're going to see	now we're going to	we're going to get
			you're going to go	I'm going to get	I'm going to ask
			you're never going to	I'm going to give	I'm going to be
					I'm going to have
		Impersonal	it's going to be	is going to be	there's going to be
			it's not going to	this is going to	are we going to
			going to be able		
		B4. Ability			
		Personal			
		Impersonal			

II.	**Discourse Organizer**				
	A	Topic Introduction	what I'd like to	if you look at	if we're going to
			and I'd like to	if you think about	let's look at the
			I'll tell you what	what we're going to	let's take a look
			I'd like to do	when you think about	it turns out that
			I'd like to show	as you can see	it turned out to
			I'd like to talk	you can see it's	if you don't know
			I'd like to show	I'm here to tell	it turns out it's
	B	Topic elaboration/ clarification	but here's the thing	what that means is	it's kind of like
			what's happening in the	what's going on in	when it comes to
			what's going to happen	so that's what I	what that means is
			to figure how to	nothing to do with	what we find is
			to figure out what	for those of you	
III.	**Referential Expressions**				
	A	Identification Focus	this is one of	one of the most	to be one of
			one of the things	for the first time	in the same way
			and this's not that	to come up with	there's a couple of
			and it's not just	the other side of	at the end of
			is that there's a		
	B	Imprecision			
	C	Specification of Attributes	there's a lot of	this is a little	
			how many of you	the rest of the	
			all of the sudden		
	D	Time/ Place / Text Reference			
		D1) Place reference	in the United States		
		D2) Time reference	at the same time	at the same time	for a long time
			in the middle of	end of the day	
			at the end of	a few years ago	
		D3) Text-deixis			
		D2) Multi-functional reference			

IV.	**Special Referential**				
			thank you very much		
			thank you so much		

4.3.1 STANCE BUNDLES

According to Biber et al., (2004), stance bundles that identify personal feelings and attitude convey two main categories: epistemic stance and attitude/modality stance. As cited in the study of Conrad and Biber (2005), epistemic bundles tend to reveal certainty/uncertainty and possibility. Conrad and Biber (2005) also claim that personal or impersonal attribute could be sub-categorized. There is little doubt that epistemic bundles in TED are highly likely to reveal information of uncertainty (personal) which attributes to the speakers who addressed themselves as 'I' or 'we'. The samples could be identified as *I don't know if, I think that's, and we don't know what* in the following samples.

*"**I don't know if** anyone's ever felt that. And I noticed pretty soon after that that all the competitors in our space had already automated my job role."*

*"What I'm suggesting is, when you connect with people around a shared interest and action, you're accustomed to serendipitous things happening into the future, **and I think that's** what we're looking at."*

*"But using some of these same methods, we can look at the NSA's data centers, and figure out, you know, **we don't know what's** going on there."*

It seems that impersonal stance bundles that show similar meaning indirectly to an individual are unlikely to be found. Moreover, the sub-category of attitude/modality stance bundles in TED tends to express speaker attitudes towards actions in the following aspects; desire identifying (*e.g. I don't want to*), obligation/directive (*e.g. don't get me wrong*), intention/prediction (*e.g. and I'm going to, it's going to be*). Referring to the study of Biberet. al., (2004, p.390), desire bundles could be identified as a framework of self-motivated wishes and desires. Furthermore,

obligation/directive bundles tend to focus on personal expression of speakers who command the listeners (you) toward instructive actions. Next, intention/prediction bundles are highly likely to express speaker's intention or future prediction. Finally, ability bundles help identify skill and tasks which listeners should achieve. It is crucial to state that bundles in TED could be only excerpted from the first three subcategories of desire bundles, directive bundles, and prediction bundles. Examples of these TED bundles could be illustrated as in the samples below.

*"**I don't want to** suggest that it's at all possible to get a rounded picture of a country simply by reading one book."*

*"All I want is for my children to be happy. And **don't get me wrong**: I think happiness is a wonderful goal for a child."*

*"The exact size and shape of these tabletops is the same, **and I'm going to** prove it to you."*

*"Now we're going to all do this together. **It's going to be** fun."*

4.3.2 DISCOURSE ORGANIZER

Referring to the study of Biber et al., (2004) and Bal (2010), the lexical bundles introduce both topic introduction and topic elaboration/clarification. As cited in the study of Conrard and Biber, (2005) there are slightly different functions where introduction bundles tend to introduce a new topic (*e.g. If you look at,*) while topic clarification is highly likely to provide more information of a previous statement (*e.g. what that means is, but here's the thing*) as in the following excerpts from TED Talks:

*"And sure enough, **if you look at the** kinds of values that come in, you see wealth, adventure, achievement, pleasure, fun, be respected, before the change, and much more post-materialist values after."*

*"What's important though is that writing only emerged about 5,000 years ago. So **what that means is** that all the people before there was any writing, every word that they ever said, every utterance disappeared."*

"Sure, I'll give you 100 reasons why coming out of my closet was harder than coming out of yours, ***but here's the thing****: Hard is not relative."*

4.3.3 REFERENTIAL EXPRESSION

There is reason to be confident that the overall bundles of referential expression in TED were found in three type categories: identification bundles (*e.g. this is one of*), specification of attributes (*e.g. there's a lot of*) as well as time and place reference (*e.g. in the middle of, at the end of*). Four-word lexical chunks of this category could be illustrated in the samples below.

"And ***this is one of*** *the early music videos that I made."*

"There's a lot of stuff we can do. ***There's a lot of*** *schools doing farm-to-school programs."*

"You mean if I want pork chops, even ***in the middle of*** *the night, your guy will fry them up?"*

"And ***at the end of*** *the day, my husband looks disappointed because I'd rather go to bed than go to the movies."*

4.3.4 SPECIAL CONVERSATIONAL FUNCTION

Finally, the special conversational functions which tend to express politeness *(e.g. thank you very much)* recorded the highest frequency in TED.

Taking into account a stereotypical oral register, Biber et. al. (2004) have revealed that the characteristics of oral register could be mainly identified in three aspects: high interaction, personal stance expression, and real time production. To put it more simply, personal stance with a high interaction of speakers is highly likely to be carried out through real time spoken utterance. Focusing through this research study, public speaking is considered as an oral register where real–time production, personal concerns and interaction are marked among participants. There is little doubt that the content and structure of speaking sessions are normally pre-planned to achieve goals and engage audiences much more effectively.

To put it more simply for functional analysis in TED, it appears that the functional type of common four-word lexical bundles in the study tends to reflect the communicative purpose of public speaking conveying and focusing on personal thought and attitude. Moreover, the speech in public speaking tends to be engaged with topic introduction and politeness concern. In terms of the highest proportion of stance expression, the speakers tended to use items relating to personal desire and attitude expressing to communicate with their audience. As cited in the study of Conrad and Biber (2005), the stance bundles are used to draw the audience into the speech presentation and activate their background knowledge before the speech begins. Next, there is little doubt that the discourse organizing from this TED corpus tended to focus on topic introduction that enabled speakers to provide a signal to audiences for the wider context of public speaking. For referential expression, the bundle identified as the referential group in the TED corpus are highly likely to reflect the experiences of the real world as well as provide information about the purpose of the speaker. Finally, due to the relatively high word frequency of special conversational functions marked as top five from the TED Corpus, the significant evidence shows that the common phrase of 'thank you very much' is widely used in public speaking to convey politeness when ending a session.

4.4 THE RELATIONSHIP BETWEEN STRUCTURE AND FUNCTION OF LEXICAL BUNDLES

There is little doubt that the relationship between structural taxonomy and discourse function of lexical bundles in TED is highly likely to be strong. Regarding the TED corpus, it seems that stance bundles, discourse organizers and referential bundles are the functional categories associating all three-structure types. However, there are slightly differences in that most stance bundles are comprised of verb phrases and organizers are mainly occupied with dependent clause bundles, while referential bundles most contained prepositional phrase or noun phrases. Finally, special referential expression is the only category associating with noun phrases. Undoubtedly, the finding patterns reveal that public speaking is highly likely to use stance bundles compiled with verb phrases. It is crucial to state that the finding in

this TED corpus reveals the interaction of structural taxonomy, functional discourse, and situational characteristics of the particular register, as shown in Figure 4.

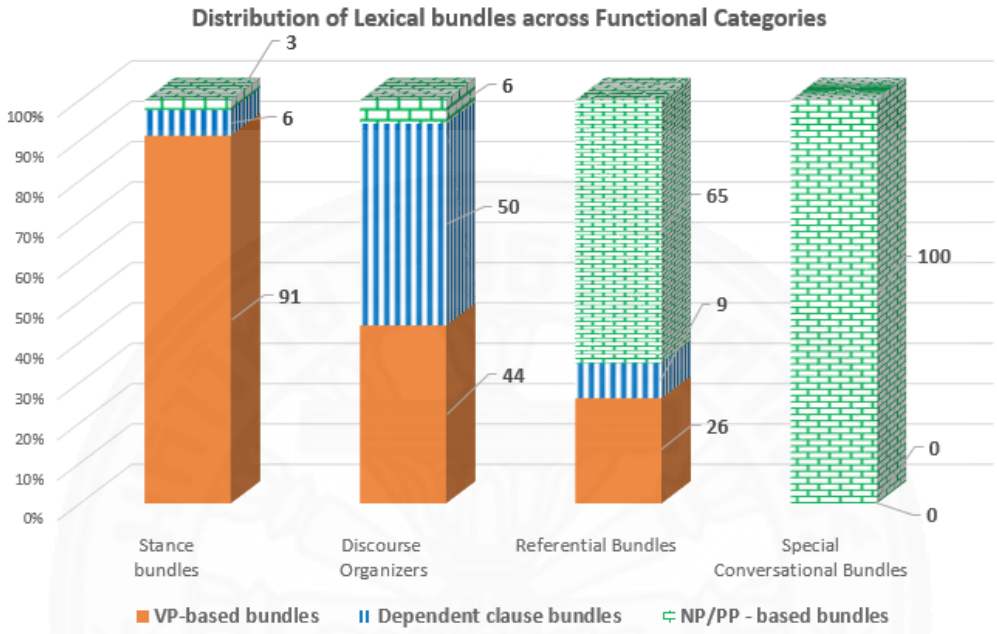


Figure4. Distribution of lexical bundles across functional categories of TED

4.5 DISCUSSION

It is crucial to state that the findings in the study are similar to the significant previous studies of Conrad and Biber (2005) conducting the research into the word frequency and use of lexical bundles in conversation and academic prose, and Biber et al., (2004) exploring the lexical bundles in university teaching and textbooks, in the following aspects. Biber et al., (2004) also claim that using nouns, verbs and personal pronouns are more common in the conversation and classroom registers, which are parallel with the results found in the TED analysis. Furthermore, the findings regarding distribution of lexical bundles across structural types in the TED corpus is similar to the conversation register conducted by Biber et al., (2004) and Conrad and Biber (2005) where a large inventory of lexical bundles in the utterances tends to have more verb phrases and dependent clause bundles rather than noun phrase and prepositional phrase bundles. In terms of functional taxonomy, the findings in the

TED corpus also correlates with the study of Biberet. al., (2004) who claimed that most bundles in the conversational register are comprised of stance bundles, discourse organizers, referential bundles and special bundles respectively. It may be surprising that the small number of the special conversational bundle "*thank you very much*" is considered as the highest frequency bundle found in the TED corpus. Of course, this implies that the public speaking register tends to have certain conventions of politeness. Obviously, in parallel with the study of Biber et al., (2004), the pattern of the interaction between form and function for lexical bundles are directly associated. Most noun and prepositional phrases tend to become referential bundles and special conversational bundles. However, the prominent finding in this study of a public speaking register is that most of the verb phrase bundles will become fixed as stance bundles. Furthermore, stance prediction bundles use all verb phrases. while the dependent clause bundles will mostly become discourse organizers. It is highly likely that the stance bundles comprised of verb phrases such as *I'm going to show*, and *we're going to see* tend to reflect the communicative purpose and convey personal thoughts, intention, prediction and desires of the speakers.

To put it into a pedagogical point of view, the findings regarding this TED corpus are highly likely to be beneficial in material design or course development for public speaking. Due to the fact that the language used in TED is real and authentic, the learners utilizing these bundles sound more natural when speaking. Moreover, samples in TED tend to deductively guide learners in how to produce language much more accurately. Taking into account the implications of lexical chunk instruction, as cited in the study of Chun-guang (2014), it is crucial to point out that lexical chunks have a great impact on language learning and teaching. Using these is considered as an effective approach in language teaching to deliver accuracy and fluency in the language production of learners (Lewis, 1993; Nattinger & DeCarrio, 1992). In terms of public speaking instruction, it is highly likely that by incorporating corpus driven analysis from the TED corpus, learners are enabled to produce their speech much more idiomatically and overcome negative L1 transfer (Chun-guang, 2014). Besides, the ready-made chunks from TED tend to enhance learners' speaking skills since they are unlikely to translate word by word from Thai to English. With functional lexical

chunk analysis from TED, teachers are able to offer guidance in public speaking structure by deploying the samples of language usage throughout the beginning, middle and ending sessions of public speaking much more effectively.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

This chapter provides a summary of the study, a summary of the findings,conclusions, and recommendations for further research.

5.1 A SUMMARY OF THE STUDY

It is fairly certain that using lexical bundles has given a new perspective to ESL/ EFL teaching and learning. In this research, the analysis of structural and functional lexical bundles used data established and retrieved from a TED corpus. Referring to the frequency based corpus driven approach, the interpretation in both structural and functional criteria brings us significant resources for English language teaching. As cited in the study of Biber et al., (2004), the pattern between form and function of lexical bundles is strongly associated in a conversational register in which the study tends to define the correlation of structural and functional discourse. To put it more simply, incomplete grammar structural chunks tend to begin with a clause or phrase such as a verb phrase, a dependent complement, or a prepositional phrase. It is highly likely that the conversational chunks provide the audience with an information framework in the aspect of functional taxonomy - stance bundles, discourse organization, or referential statuses (Biber et al., 2004).

The results of this study provide new insights into the development of language teaching going beyond the expectation of target language usage. It could be said that with an effective corpus driven approach, a public speaking course is better designed to meet specific needs of learners efficiently. There is little doubt that beneficial communicative models from the TED corpus can not only motivate Thai EFL language learners but also help enhance accuracy and fluency in public speaking for a prosperous career path.

5.2 IMPLICATIONS

There is little doubt that a corpus driven approach helps facilitate teachers to have a better understanding of which vocabulary - as lexical chunks - learners should attain. As a result of this, the appropriate material design should be applied effectively to ensure that learners are enabled to develop the awareness of lexical use for their utterances much more effectively (Lewis, 1997). Hence, it is very important, and more interesting for teachers, to implement and investigate how corpus driven lexical chunk instruction as a lexical approach helps enhance language learners' output in respect to public speaking. Besides, in order to ensure that learners are enabled to use lexical chunks in their speech properly, language teaching should be contextualized throughout the context (Chun-guang, 2014). It is fairly certain that without the contextualization of presenting TED Talks as a public speaking model, Thai learners are likely to be demotivated. For this reason, a sample lesson plan using TED Talks is provided in the appendix below.

As cited in the study of Lewis (1997): "Language is acquired by understanding messages" (Krashen and Terrel, 1983). However, Lewis (1997) has claimed that in order to process input for language acquisition, the activities that raise conscious awareness of learners should be prominent and make use of lexical chunks to produce the language naturally. The work of Bareggi (2006) has indicated that a lexical approach could be identified as a "language acquisition process or language teaching approach that enables learners to be able to put proper words in a proper place (p.2)." Moreover, Bareggi (2006) also points out that the essential principal of a lexical approach could be identified as consisting of grammatical lexis and multi-word chunks, not lexicalized grammar or single word vocabulary, developing students' awareness in language teaching, structuring grammar as subordinate to lexis, focusing on fluency rather than accuracy of language usage (Lewis, 1993). Analysis of word frequency occurrence by using multi–words as a set of lexical chunks helps enhance learners to use them correctly and enables them to produce natural language successfully (Bareggi, 2006).

As Lewis (1997) has claimed, individual words and traditional grammar structures tend to be less focused on within the effective lexical approach in which lexical items carrying the meaning are emphasized for substantial meaning of both written and spoken context.

5.3 RECOMMENDATIONS FOR FURTHER RESEARCH

In the present study, the research on the topic is still in its beginning stage in which some limitations are highly likely to exist. First of all, due to the time limitation, the collected corpus size was approximately 500,000 words, which is unlikely to be fully representative for language learners. As a result of this, a bigger corpus size of TED should be processed for more reliable conclusions. Additionally, it would be useful to investigate more materials used for public speaking to see if these frequently used four-word lexical bundles do appear for further effective material development. Furthermore, it would be more interesting if the lexical chunks could be considered on a particular theme, which would be beneficial for further study of English for specific purposes. Finally, in respect of productive skills which tend to be the most difficult parts for non- native learners, a study of lexical bundles used in TED could be conducted to compare with the lexical bundles used in writing. This comparison could help enhance productive skills in English much more effectively.

REFERENCES

AntConc Corpus analysis software. Retrieved May 10, 2016, from http://www.laurenceanthony.net/software.html.

Bal, B. (2010). Analysis of four-word lexical bundles in published research articles written by Turkish scholars. *Applied Linguistics and English as a Second Language Theses*, Georgia State University, United States.

Bareggi, C. (2006). The lexical approach. *Language matters, 13,* 2-10.

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

Boonkit, K. (2010). Enhancing the development of speaking skills for non-native speakers of English. *Procedia-Social and Behavioural Sciences, 2,*1305-1309.

Chun-guang, T. (2014).An empirical research on the 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. Portland State University.

Hamilton, C., & Parker, C. (1996). Communicating for results: A guide for business & the professions (5thed.). Belmont, CA: Wadsworth Publishing Company.

Jablonkai, R. (2009). "In the light of" : a corpus-based analysis of lexical bundles in two EU- related registers. *Corvinus University of Budapest: WopaLP, 3,* 1-26.

Jidathai, S. (2015). Factors affecting English speaking problems among engineering students at Thai-Nichi Institute of Technology. *The 3rd National Interdisciplinary Academic Conference (TNIAC 2015),* 344-348.

Jukneviciene, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. Vilnius University.

Kunnu, W.,& Sukwises, A. (2014).Teaching speaking skills to adult language learners through ALM. *International Journal of Social, Behavioral, Educational, Business and Industrial Engineering*, *8*(8), 2715-2717.

Lackman, K. (2012). Lexical approach activities: A revolutionary way of teaching. *Ken Lackman & Associates: Educational Consultant*. Retrieved from https://www.yumpu.com/en/document/view/6731065/lexical-approach-activities-ken-lackman.

Lewis, M. (1997). *Second language vocabulary acquisition*. Cambridge: University Press.

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

Plangkham, B.,& Porkaew, K. (2012). Anxiety in English public speaking classes among Thai EFL undergraduate students. *Learn Journal: Language Education and Acquisition Research Network*, 110-119.

Randall, M. (2007). *Memory, psychology, and second language learning.* Amsterdam/ Philadelphia: John Benjamins.

TED Talks. Retrieved from https://www.ted.com/about/our-organization.

Wang, Y. C. (2012). An exploration of vocabulary knowledge in English short talks: A corpus - driven approach. *International Journal of English Language Teaching*, *2*(4), 33-43.

Yusu, X. (2014). On the application of corpus of contemporary American English in vocabulary instruction. *Canadian Center of Science and Education, 7*(8), 68-73.

Zhao, Y. (2009). An empirical study of L2 learners' use of lexical chunks and language production. *The School of Teacher Education.* Kristianstad University College, Sweden.

APPENDIX

APPENDIX A: LESSON PLAN

Topic: Useful Topic for Public Speaking [Beginning Session of Public Speaking]

Duration: 90 minutes

Number of students: 20 students

Proficiency level: Upper Intermediate (Graduated Employee: Age 24-35)

Objective

1. To develop students' awareness of lexical chunk usage in public speaking.
2. To expand students' knowledge of vocabulary usage as chunks in public speaking.
3. To develop speaking skills/discussing preferences and attitude towards public speaking by using stance expression.
4. To promote individual and cooperative learning.

SWABATS

1. To be able to notice and be aware of lexical chunks in public speaking.
2. To be able to use vocabulary as lexical chunks, stance expression for public speaking correctly.
3. To be able to discuss preference, attitude and give a short speech for public speaking fluently.
4. To be able to perform as individual and group.

Previous Class work

1. Introduction to public speaking.

Time (min)	Activities	Procedure		Goals	Materials
		Teacher	Students		
10 mins	Warm up	1. T shows a clip of Ted Talk " Living the dream [Praya Nataya Lundberg] TEDxYouth@NIST" (ปู ไปรยา ลุนด์เบิร์ก)	1. Ss respond to the clip.	1. To activate the Ss' schemata of public speaking.	https://www.youtube.com/watch?v=m1LeGwELv9o
		2. T ask Ss to discuss about the topic of Praya's public speaking.	2. Ss brainstorm and discuss about the topic.	2. To bring more about the topic in public speaking.	
25 mins	Presentation	1. Group work:T shows Ss the transcripts of Praya's public speaking. Then assign the comprehension questions and discuss whether she is an effective speaker.	1. Ss respond to the questions.	1. . To raise the awareness of effective speech e.g. body language, language usage, suitable expression.	Transcript of Praya's public speaking
		2. T assigns Ss to read the transcripts again then read and look for the expression (stance expression) that are used for the talk.	2. Ss work as group and respond to the task.	2. To develop Ss' noticing skill of lexical bundles used in public speaking.	Transcript of Praya's public speaking
30 mins	Practice	1. T elicits the stance expression and explain the meaning where necessary. Then provides slot fillers for expression. For example; * I'm going to … * We're going to … * I don't know ….	1. Ss drill for the target stance expressions.	1. To get Ss to practice using stance expression for public speaking.	Flash card of sentence stance
25 mins	Production	**Free Practice :** 1. T asks Ss to work in pairs and discuss the preferences topic for public speaking. Then give a short talk in front of the class for 5 mins using stance expression.	1. Ss repose to the tasks.	1. To develop public speaking skills of being able to choose the useful topic and use the proper sentence stances focusing on the beginning session of the talk.	

BIOGRAPHY

Name	Ms. Nutchada Suwanwong
Date of Birth	October 25, 1982
Educational Attainment	2004: Bachelor of Arts Faculty of Arts, Chulalongkorn University
	2009 : MSC Digital Marketing Huddersfield University (United Kingdom)
Work Position	January 2016 to present English Teacher : Inspire English Institute, Bangkok
Work Experiences	March 2014 – October 2015 English Teacher (Part Time): ESL School, Bangkok March 2012 – October 2015 Marketing Communication : QI Services, Bangkok March 2010 – January 2012 PR Officer : Noble Development, Bangkok