

**AN ANALYSIS OF SPOKEN TRANSCRIPTS IN ENGLISH UNIVERSITY
TEXTBOOKS: FORMULAIC SEQUENCE OCCURRENCE**

A MASTER'S PROJECT

BY

JOBETH A. JANG

**Presented in Partial Fulfillment of the Requirements for the Master
of Arts Degree in Teaching English as a Foreign Language
at Srinakharinwirot University**

January 27, 2008

**AN ANALYSIS OF SPOKEN TRANSCRIPTS IN ENGLISH UNIVERSITY
TEXTBOOKS: FORMULAIC SEQUENCE OCCURRENCE**

AN ABSTRACT

By

Jobeth A. Jang

**Presented in partial fulfillment of the requirements for the Master of
Arts degree in Teaching English as a Foreign Language (TEFL)
at Srinakharinwirot University**

January 27, 2008

Jobeth A. Jang. (2008). *An Analysis of Spoken Transcripts in English*

University Textbooks: Formulaic Sequence Occurrence. A Master's Project, M.A. (TEFL). Bangkok. Graduate School, Srinakharinwirot University. Project Advisor: Dr. Saengchan Hemchua

Linguistic observation of discourse and lexical words has shown natural and authentic use of language and the existence of formulaic sequences. This study is concerned with the occurrence of formulaic sequences in English university textbooks to indicate that vocabulary teaching need not always be focused on individual words. It analyzed and classified formulaic sequences into two main categories: polywords and multi-word expressions. The transcripts of three intermediate level textbooks were compiled into a spoken corpus consisting of 19,519 running words. A concordance program was used to examine the frequently occurring formulaic sequences in the corpus. It was found that the total number of formulaic sequences in the spoken corpus was 565: 428 multi- word expressions and 137 polywords. It indicated a high frequency of occurrence in multi-word expressions. Based on each of the categories, fixed expressions appeared with the highest occurrence while idioms appeared the lowest. The results suggested that in the English university textbooks transcripts, conversation exercises for listening and speaking skills use multi-word expressions extensively. An implication in vocabulary teaching is that learners should be provided with opportunities in utilizing formulaic sequences specifically multi-word expressions in classroom instruction and activities in order to produce the language in a manner similar to native speakers.

การวิเคราะห์บทสนทนาในแบบเรียนภาษาอังกฤษระดับมหาวิทยาลัย : Formulaic Sequences

บทคัดย่อ

ของ

โจเบต จัง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต สาขาวิชาการสอนภาษาอังกฤษ
ในฐานะภาษาต่างประเทศ

มกราคม 2551

โจเบต จัง. (2551). การวิเคราะห์บทสนทนาในแบบเรียนภาษาอังกฤษระดับมหาวิทยาลัย :

Formulaic Sequences. สารนิพนธ์ ศศ.ม.(การสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อาจารย์ที่ปรึกษา
สารนิพนธ์: อาจารย์ ดร. แสงจันทร์ เหมเชื้อ.

งานวิจัยนี้มุ่งศึกษาสำนวน (Formulaic Sequences) ที่ปรากฏในบทสนทนาในแบบเรียนภาษาอังกฤษระดับมหาวิทยาลัย โดยใช้โปรแกรมสำเร็จรูปในการหาความถี่ของสำนวนที่ใช้ ซึ่งแบ่งเป็น 2 ประเภท คือ Polywords และ Multi-word expressions จากการวิเคราะห์พบว่า สำนวนที่ใช้มากคือ Multi-word expressions (428 สำนวน) ส่วนสำนวนแบบ Polywords มีการใช้น้อยกว่า (137 สำนวน) และ Fixed expressions มีการใช้ในบทสนทนามากที่สุด ส่วนสำนวนแบบ Idioms มีการใช้น้อยที่สุด ทั้งนี้ผลการวิเคราะห์สามารถนำไปประยุกต์ใช้ในการเรียนการสอนศัพท์ โดยมุ่งเน้นการสอนสำนวนมากกว่าการสอนคำเดี่ยว เพื่อให้ผู้เรียนสามารถพูดภาษาอังกฤษได้อย่างเจ้าของภาษา

**AN ANALYSIS OF SPOKEN TRANSCRIPTS IN ENGLISH UNIVERSITY
TEXTBOOKS: FORMULAIC SEQUENCE OCCURRENCE**

A MASTER'S PROJECT

By

Jobeth A. Jang

**Presented in partial fulfillment of the requirements for the Master of
Arts degree in Teaching English as a Foreign Language (TEFL)
at Srinakharinwirot University**

January 27, 2008

Copyright 2008 by Srinakharinwirot University

The Master's Project Committee and Oral Defense Committee have approved this Master's Project as partial fulfillment of the requirements of the Master of Arts degree in Teaching English as a Foreign Language (TEFL) of Srinakharinwirot University.

Project Advisor

.....

(Dr. Saengchan Hemchua)

Chair of the Master Degree in Teaching English as a Foreign Language (TEFL)

..... TEFL Chair

(Dr. Wanlapa Thaijinda)

Oral Defense Committee

..... Chair

(Dr. Wanlapa Thaijinda)

..... Committee

(Dr. Aurapan Weerawong)

The Faculty of Humanities has approved this master's project as partial fulfillment of the requirements for the Master of Arts degree in Teaching English as a Foreign Language (TEFL) of Srinakharinwirot University.

..... Dean of the Faculty of Humanities

(Associate Professor Chaleosri Pibulchol)

January 27 , 2008

ACKNOWLEDGEMENTS

I wish to express my utmost appreciation to Dr. Saengchan Hemchua, my Project Advisor, for her compassionate guidance, generosity and helpful resources throughout this study. I am grateful to the Committee of Oral Defense Dr. Aurapan Weerawong, Dr. U-maporn Kardkarnklai and Dr. Wanlapa Thaijinda for their inquisitive comments and valuable suggestions.

I would like to extend my gratitude to all the Great Instructors of the TEFL program for imparting their knowledge and skills to all the participants, and giving their loving guidance for this study.

I would like also to extend my thanks to my friends and colleagues at Rajamangala University of Technology Thanyaburi for their concern and sufficient consideration for this project.

Special thanks to my loving husband, Soung Gun Jang for his consistent support, constructive criticisms and practical suggestions.

Above all, I would like to dedicate this endeavor to my Savior, Jesus Christ for His Divine Guidance and Provision.

Jobeth A. Jang

TABLE OF CONTENTS

CHAPTER	PAGE
1 INTRODUCTION.....	1
Objective of the Study	5
Significance of the Study.....	5
Scope of the Study.....	6
Definition of Terms.....	6
2 REVIEW OF RELATED LITERATURE	10
Formulaic Sequences.....	10
Second Language Vocabulary Acquisition.....	17
Spoken Discourse Production	21
Previous Related Studies	23
3 METHODOLOGY	26
Data Collection	26
Research Instruments	27
Data Analysis	31
4 FINDINGS	33

TABLE OF CONTENTS (CONTINUED)

CHAPTER	PAGE
5 CONCLUSION AND DISCUSSION.....	47
Pedagogic Implications.....	51
Limitations of the Study	53
Suggestions for Further Studies.....	53
 BIBLIOGRAPHY	 55
 APPENDICES	 62
Appendix A Transcript Samples	62
Appendix B Sample Analysis	63
Appendix C Sample of Polywords	64
Appendix D Sample of Multi-word Expressions	65
 VITAE	 67

LIST OF TABLES

TABLE	PAGE
1 Formulaic sequence occurrence in the three English language university textbooks	32
2 Frequency of polyword occurrence in the three English language university textbooks.....	34
3 Frequency of multi-word expression occurrence in the three English language university textbooks.....	35
4 Ten high frequency words in the three English language university textbooks	36
5 Sample concordance of collocations	37
6 Sample concordance of binomials	38
7 Sample concordance of compounds	39
8 Sample concordance of phrasal verbs.....	40
9 Sample concordance of prepositional phrases.....	41
10 Sample concordance of semi-fixed expressions.....	42
11 Sample concordance of fixed expressions	43

LIST OF FIGURES

FIGURE	PAGE
1 The elements of pedagogy in conversation	18
2 Frequency and percentage of the clusters of <i>as</i>	25
3 Concordance lines of the preposition <i>to</i>	26
4 Formulaic sequences in the three English university textbooks.....	33

CHAPTER 1

INTRODUCTION

In recent years, numerous sets of multiword units which are arbitrarily use in native speakers' discourse known as formulaic sequences have taken a considerable interest in research and second language learning and teaching. It has been found that speakers and writers alike use strings of words that are readily accessible for communication. This is because speakers and writers produce language in an uninterrupted manner using formulaic sequences. On the other hand, listeners and readers do not always interpret or connect word by word but rather make use of formulaic sequences for comprehension (Howarth.1998).

In second language classrooms, learners use formulaic sequences either as a means to carry on a fundamental conversation or in a quest of the learners' desires to produce correct forms (Bygate.1988). Formulaic sequences also play a significant role, not simply as a linguistic unit, but rather a tool that can be put to many uses such as to maintain communicative production, aid in the development of learning strategies, and a naturalistic use for classroom learning (Weinert.1995). Therefore, the awareness of the frequent occurrence of formulaic sequences in native speakers' discourse is important for second language learners for a reason that they can creatively make use of similar sequences in producing the language either in classroom contexts or real situations. Having much appeal in second language pedagogy, the analysis on formulaic sequences

has attracted other language related disciplines such as corpus linguistics, discourse analysis, language processing, and second language acquisition research.

Traditionally, the identification of formulaic sequences is once a taxing work, however, recently with the aid of corpora the output of formulaic sequences in the native speakers' discourse has been profoundly discovered. These corpora contain the countless linguistic production of native-speakers in different contexts. They contain the data that display the manifestation of linguistic mental processing. Practically, once a corpus is set, a program is used to analyze words in association with their grammatical and lexical components for example sorting high frequency, concordance, and citations of word phrases. From the data in the corpus, a researcher or reader can also differentiate the words and/or phrases in relation to how much knowledge speakers assume, what kind of relationship they have, and what context they are in (Sinclair. 1991). In other words, corpus analysis can be used to describe more complex multi-word combinations and phrases. For a language teacher, a corpus can be used to identify the most exemplars to teach. As a result, corpus-based investigation of word combinations provides linguists with a detailed and clearer picture of the language in a way that it provides a natural approach of the language (Schmitt. 2002; Stubbs. 2002). Thus, the use of corpora in the analysis of formulaic sequences benefits both researchers and language teachers because they reveal the various delicate compositions of native speakers' language.

Pedagogically, formulaic sequences should be included in the teaching syllabus as a learning input. In the classroom, language teachers should consider activities demonstrating multi-word patterns showing their usage and lexical functions. In addition,

although English is accepted as a foreign language in Thailand, there is a scarcity in native language teachers; apparently, most English language classes are likely to be taught by non-native language teachers. Therefore, including formulaic sequences in teaching materials and classroom activities would provide non-native language teachers with possibilities to model the language in a natural way, essentially, with the aim to attain higher proficiency in second language learners' fluency.

Furthermore, one of the main goals in second language teaching is to prepare learners to independently produce and creatively use the language. In view of this, there is a consensus that mastery of formulaic sequences equates successful language learning and native speaker fluency (Lewis. 1997; Nattinger & DeCarrico.1992; Stubbs. 2000). However, it is found that language learners have difficulty in mastering formulaic sequences (Moon.1998; Yorio.1980). When learners produce structurally correct language, it does not sometimes sound natural. They tend to overuse certain relatively infrequent words and under use certain relatively frequent words. There are two possible causes that hinder learners from obtaining the goal (Wray.1999). One possible cause that second language learners sound non- native is a lack of direct language experience. Although formulaic sequences are commonly used in media for instance entertainment and modern communication, learners could hardly notice the sequences because there is insufficient interaction for the learners to practice the language. A lesser interaction, therefore, limits learners' language production. Wray (2000) states input without adequate interaction is not satisfactory for language acquisition. Another probable cause is that formulaic sequences are not properly taught. In second language classrooms, the

language focus does not match the real world experience of language. Thus, learners appear non-native by virtue of having too little of the fairly wide and frequent words, and too many of the fairly rare. Therefore, for second language learners to sound native-like they should be led in the observation and imitation of formulaic sequences that are recurrent in a given discourse.

Consequently, recent attempts have been made to introduce formulaic sequences in second language teaching. For example, Willis (1990), in designing syllabus, favors the procedures of prominent patterns believing that learners need to notice patterns and speculate them in their learning. Nattinger & DeCarrico (1992) focus on the usefulness of formulaic sequences in teaching conversation. Meanwhile, Lewis (1997), in teaching, claims that language is comprised of multiple patterns of lexical phrases not grammaticalized individual words. Thus, he posits that language should be taught in a larger vocabulary especially the pragmatically useful lexical items, lexical approach. Lexical approach focuses on exposing second language learners through exposure, instruction, and practice on the language patterns; eventually, in the learners' advance stage, grammar rules should be used to analyze the recognized items.

Apart from the recent approaches in teaching formulaic sequences, numerous research studies were also made to analyze formulaic sequences in native spoken language and written texts in various fields such as studies of formulaic sequences in writing (Jones & Haywood. 2004; Lin. 2003; Oakey. 2000) and collocations in textbooks (Chamnan. 2004; Koya. 2004). Although it is evident that much work has been done, scant attention has been paid to analyze the occurrence of formulaic sequences in spoken

transcripts in English textbooks. As known, besides language teachers, textbooks are a rich source of input in learning a language since they provide a fair degree of authentic samples of English that are vital for university students. McCarthy and Carter (1997:38) state that the spoken language is the main source of exposure to language. In second language classrooms, textbooks are most likely the main source of teachers' output and learners' input. Therefore, this study focuses on formulaic sequences in transcripts for the reason that transcripts are crucial components in teaching and practicing particularly listening and speaking skills. To a second language teacher, formulaic sequences are a handheld device for teaching students while to a second language learner, they are valuable practical guides in the absorption of a native – like language.

Objective of the Study

This study aimed to examine the feasibility formulaic sequences worth teaching to second language learners. It analyzed those sequences that occur in the transcripts of English university textbooks.

Research Questions

This study attempted to answer the following research questions.

1. What are the formulaic sequences found in the transcripts of the target English university textbooks?
2. How frequently do the formulaic sequences occur in the target English university textbooks?

Significance of the Study

This study would be useful to general audience, specifically second language teachers and learners. For second language teachers, they can utilize formulaic sequences as a tool in teaching vocabulary in a broader sense. For second language learners, they can use the knowledge of formulaic sequences to generate their comprehension and communication. In addition, the findings would be useful to researchers who want to do further research on this topic or extend this knowledge to any other related fields of study. Finally, for textbook writers it would serve as a guiding principle in the development of teaching materials necessary for successful second language teaching and learning.

Scope of the Study

The study covered the following extent.

1. This research examined the formulaic sequences occurring in the transcripts of three English textbooks: *English Upgrade 2* (Gershon, et.al., 2002), *Interchange 2* (Richards, et.al., 2005) and *Person to Person 2* (Richards, et.al., 2005).
2. The three English textbooks are used mainly at intermediate level for university students.
3. The samples of the study comprised of 29, 30 and 32 transcripts in each textbook respectively: a total of 91 transcripts for three English textbooks.
4. Pronunciation practice and listening answer keys were excluded.

Definition of Terms

For clarity of the terminology, the terms used in the context of this study are defined as follows.

1. **Spoken discourse** refers to the language produced during a real time conversation the speaker and listener interactively exchange information to get their message delivered.
2. **Formulaic sequences (FSs)** refer to sets of prefabricated strings of words in spoken texts that have pragmatic meaning but function as one unit. For example, *By the way* is a sequence that functions to signal a statement; *Many happy returns* is a sequence that functions to wish someone well; and *How are you doing?* is a sequence that functions to express sympathy.
3. **Multi- word expressions** refers to the phrases or sentences that have fixed or semi-fixed composition. In this study, they are grouped as prepositional phrases such as *in the bottom shelf*, idioms such as *eyes are bigger than one's stomach* meaning *someone wants to have more food than he/she can eat*, fixed expressions such as *I think so*, and semi-fixed expressions such as *Where is the way to ____?*
4. **Polywords** refer to the two or three words that have fixed composition in a native speech but have single meaning. In this study, they are grouped as binomials such as *bread and butter*, compounds such as *white house*, collocations such as *kind heart* but not *unkind heart*, and phrasal verbs such as *head up* meaning *be careful*.
5. **Corpus** refers to a computerized collection of texts from a large number of

written and spoken languages that are installed, and readily accessed for analysis. In this study a collection of the transcripts from three English university textbooks, a spoken corpus, consist of at 19,519 running words.

6. **Concordance** refers to a comprehensive listing of a given item in a corpus, most often a word or a phrase, also showing its immediate context. (Hunston,2002), as in the following sample of the word *blessing* adapted from the *Bible Text Easy English* version:

315 on their father, and do not give a blessing to their mother. There is a gener
316 place has been named the Valley of Blessing to this day. Then all the men
317 on you and will put a curse on your blessing: truly, even now I have put a curs
318 old ring. Job 42:12 And the Lord's blessing was greater on the end of Job's li
319 They may give curses but you give blessing; when they come up against me, pu
320 all generations give witness to the blessing which has come to me. Luk 1:49
321 give blessing; and he has given a blessing which I have no power to take awa
322 hildren of Israel. Now this is the blessing which Moses, the man of God, gave
323 he is able, in the measure of the blessing which the Lord your God has given
324 ill be cursed by the people; but a blessing will be on the head of him who lets
325 for two hundred and ninety days. A blessing will be on the man who goes on wai
326 qually from mind.For if you give a blessing with the spirit, how will the man
327 daughters saw her, and gave her a blessing; yes, the queens and the servant-
328 did to his father, Is that the only blessing you have, my father? give a bless
329 m: have no fear for I am with you, blessing you, and your seed will be increas
330 ve ear to Balaam; and so he went on blessing you; and I kept you safe from him

7. **English University Textbooks** refer to three textbooks namely: *English Upgrade* 2 (Gershon, et.al., 2002), *Interchange 2* (Richards, et.al., 2005), and *Person to Person 2* (Richards, et.al., 2005) that are used for intermediate level learners.

8. **Transcripts** refer to the written text of conversation exercises for listening and speaking practice transcribed by material developers of the target English textbooks. (See samples in *Appendix A*).

9. **High frequency words** refer to words that occur frequently when sorted through a computer program, for example, sorting the word frequency of the transcripts in *Interchange 2* in decreasing order.

429 the	324 to	281 a	262 I
229 you	200 and	147 of	138 in
129 it	121 what	116 is	102 what
88 about	86 woman	82 for	82 man
79 have	74 listen	74 well	70 do

10. **Content words** refer to words under nouns, verbs, adjectives, and adverbs that carry much information load in a text.

11. **Function words** refer to words under articles, pronouns, prepositions, interjections and conjunctions.

CHAPTER 2

REVIEW OF RELATED LITERATURE

This chapter reviews the relevant literature including formulaic sequences, second language vocabulary acquisition and teaching, language production particularly speaking, and related literature. It also includes previous related studies.

Formulaic Sequences (FSs)

This section reviews the key features and functions of formulaic sequences including their linguistic characteristics. Having frequent instinctive occurrence and prefabricated in nature, formulaic sequences appear to be regular in their forms and meanings. Wray (2000) lucidly describes formulaic sequences

are sequences, continuous or discontinuous, of words or other meaning elements, which are, or appear to be, prefabricated that are, stored and retrieve whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.

Although formulaic sequences are prefabricated in nature, they include tightly idiomatic and unchallengeable patterns, such as *Long live the King*, which are unintelligible in meaning and grammatically irregular, and, at the other, transparent and flexible patterns containing slots for open class items such as *have a wonderful evening*; *wonderful* can be substituted with *pleasant* or *nice*. Considering the wide range of multi-word patterns from deliberately fixed to flexible patterns, Moon (1997) coins three distinctive structural features that significantly encompass the attributes of FSs. First, institutionalization, it is the process wherein the sequences become conventional in a

contextual interaction until they conform as standard. When speakers often use the same words in the same pattern in a casual conversation, storytelling, or news reporting, formulaic sequences occur solely because of frequency. As a result, the repetitive use of words and multi-word patterns in a speech, the frequently occurring patterns tend to be accepted in a certain conversational situation.

Second, fixedness, it is the degree to which a multi-word item is frozen as a sequence of words. The component words can be inflected in a predictable or regular way or from less variable to more variable, for example *a year ago*, a variable sequence can be *a day ago*, *a short while ago*, *a very long time ago*, to indicate a specific time in the past. Furthermore, fixedness transpires in memorized unanalyzed patterns when a speaker has stored units of language of clause length or longer stored as sequences in the memory for example, *Can I help you?* for offering assistance and for greeting *It's really great to see you!* Likewise, the conditional marker *if I have time, I will show you*, the syntactic frame is *if + I + simple present phrase, I + simple future phrase* can be combined freely using other component words.

The third feature is that of non-compositionality, it is the degree that a multi-word item cannot be interpreted on a word-by-word basis but a holistic unit with single meaning such as word combinations like idioms. For example *follow on one's footsteps* means *to follow one's example* and a phrasal verb for example *get along* meaning *to be on good relation with someone socially*. These distinguished features of formulaic sequences being conventional, fixed or semi-fixed, having syntactic or semantic meanings operate together in spoken language in conjunction with their phonological

criteria. In essence these features, also, are the most probable fundamental basis in classifying formulaic sequences.

Aside from the distinct features of formulaic sequences, there are three different models identified to categorize the phenomena. The categorization starts with the semantic based model, the most traditional. It attempts to differentiate types of formulaic sequences according to the degree of compositionality. Most likely the sequences coincide with idioms and metaphors (Yorio, 1980). For example *straight from horse's mouth*, and *to pull someone's legs* when used mainly as phrases without their pragmatic or direct explicational context would be vague for a hearer to understand most specially for the first time. It is, also, common for a formulaic sequence to carry a metaphorical meaning that can be retrieved directly from the mental lexicon. For example *the apple of one's eye*, and *I can read you like a book*. However, the scope of formulaic sequences is much wider than idioms and metaphor so in contrast with the semantic based, the syntax-based model is formulated. It takes the most significant feature, fixedness in the combination of sequences and differentiates either the sequences have idiomatic or non-idiomatic characteristics.

Equally, Howarth (1997) divides word combinations into two major categories; functional expressions and composite units. Functional expressions are identified by the role in discourse such as openers (*To begin with*), slogans (*Keep our city clean*), and proverbs (*A friend in need is a friend indeed*). Composite units, on the other hand, are realizations of phrase structures into their lexical and grammatical collocations. For instance, the phrase *blow a trumpet* has a lexical component *carry a trumpet*. Lexical

collocations consist of two open class items while grammatical collocations consist of one open and one close class item such as *by the time* and *at the time*. *By* and *at* are probable options to complete the close phrase *the time*. Hence, syntax based and semantic based models are similar in their scope of formulaic sequences, the inflectional and transformational manipulation of idioms. Eventually, researchers come up with another model; the latest, the function-based model. It identifies that FSs are integrated into vocabulary in terms of their pragmatics.

To an extent, benefiting speakers in every day language processing in both production and comprehension, Wray (2002) posits two categories of formulaic sequences based on their functions. FSs serve as a tool for social interaction due to the fact that speakers use ready-made word clusters to survive and to manipulate their surroundings. For example: *Keep on the right* (in giving commands)

Would you mind repeating that? (for requesting)

Another function is that FSs act as compensatory devices for short-cut processing. Since various word strings are commonly used and produced often, it is not necessary for speakers to start from scratch to produce a new language. They associate familiar word strings with new information and create a new task. Ellis (1997) shows that once the brain is familiar with a linguistic task it bypasses the processing route to learn the same task. Therefore, FSs act as assistive devices to second language learners who could manipulate multi-word patterns to establish a spoken communication in a minimized amount of processing.

In sum, the inherent features of formulaic sequences; institutionalization, fixedness, and non-compositionality are illocutionary force in their categorization. The developmental model of categorizing formulaic sequences is also a crucial element in second language teaching and learning due to the fact they would provide a closer look on how native speakers' use and produce the language. Thus, identifying the formulaic sequences with their corresponding categories would perform as guidelines for material developers and language teachers, as well as expanding language-learning strategies.

In this study, the categories of formulaic sequences are adapted from Lewis's and Moon's taxonomy. They are chosen for reason that each researcher situates categories that are complementary to the three models, in this way at least avoiding bias in the categorization of the sequences. Therefore, combining the two categories would at least cover all types of recurrent formulaic sequences. Obviously, FSs are categorized based on their structural features and their semantic, syntactic and functional characteristics. Practically, Lewis's (1997) taxonomy includes four categories namely: polywords, collocations, institutionalized utterances and sentences frames.

First, polywords are word combinations for example *upside down* and *white house*. Second, collocations referring to word combinations that co-occur frequently for example *absolutely convinced*, *clear conscience* and *community service*. Third, institutionalized utterances are sets of words that occur in a pattern conveying a single meaning for example *I'll get it*; and *That'll do*. Finally, sentence frames and heads including spoken frames refer to discourse markers or fillers in a spoken language for example *The*

fact/suggestion/ problem/ danger was...;In this presentation we will explore..., first... second...finally... Lewis's taxonomy involves both lexical and interactive expressions in the classification that is slightly comparable to the subsequent taxonomy.

Similarly, Moon's (1997) taxonomy on multi-word units includes five categories namely; compounds, phrasal verbs, idioms, fixed phrases and pre-fabricated routines. First, compounds are fashioned when two or more words are combined to make a single lexeme. This lexeme can be written either fused, hyphenated or separated such as *multiword or multi-word, wildflower, blue-eyed, and Board Member*. Second, phrasal verbs are combinations of monosyllabic verbs for example *take, stop, and give* with adverbial or prepositional particles for example *up, off, and away*. Sometimes, phrasal verbs are comprehensible from their components for example *stop over* and *give away*, but most of the time their meaning is not interpretable such as *figure out* which means to find a solution to a problem and *make up* which means to invent a story. Third, idioms are phrases with holistic meanings such as *from the bottom of my heart* or *feeling bit under the weather*. Fourth, fixed phrases are sets of words that are sternly fixed and institutionalized in a particular situation such as *Praise the Lord* a response to a minister in a church service, *have a safe journey home* to wish someone a safe trip or *it's a small world, isn't it?* when somebody unexpectedly meet someone he/she knows. Lastly, prefabricated routines are phrases, stereotyped collocations or semi-fixed strings that are use in discoursal situations as structuring devises such as *my point is...* and *I highly recommend...*As a result, combining the preceding two taxonomies and collapsing their

components would at least cover the frequently occurring formulaic sequences in the spoken transcripts.

Concurrently, it is the aim of this study to identify and analyze the occurrences of formulaic sequences, therefore, to obtain the objective, formulaic sequences are classified into two main categories: polywords refers to two or three words that have fixed composition in a native speech but have single meaning. They are grouped in the following sub-categories binomials such as *bread and butter*, compounds such as *white house*, collocations such as *kind heart* but not *unkind heart*, and phrasal verbs such as *head up* meaning *be careful* or *cheer up*. Another category is multi-word expressions referring to phrases or sentences that have fixed or semi-fixed composition with the following sub- categories; prepositional phrases such as *top of the table* and *on the corner of...*; fixed expressions such as *I think so* and *Neither do I*; semi-fixed expression such as *It's _____ than I thought*, and *That's interesting/ terrible*, and idioms such as *lend someone a hand* and *put one's self into someone's shoes*.

Evidently, the category on words is excluded from the former. For concern of ambiguity, institutionalized utterances and sentences frames were incorporated with the latter, fixed phrases and prefabricated routines. Later, categorized into semi-fixed and fixed expressions, binomials as well as prepositional phrases are included. Binomials technically, contain more than two words however with the connectors used the other two words can be easily identified. In the case of prepositional phrases, they are dealt with as a single separate group. It is perceived that these categories are commonly used in the conversation dialogs of the English university textbooks. Since the study focuses on

formulaic sequences in spoken discourse, there is a tendency for words and phrases to have repetitive occurrence. Moreover, with the many recurrent multi –word patterns, it is explicable that one problem with the study of formulaic sequences is the discovery on where the dividing line is. Therefore, to lessen confusion, the eight categories would at least provide a discrete separation among polywords and multi-word expressions.

Second Language Vocabulary Acquisition and Teaching

In this section, second language vocabulary acquisition and teaching are discussed. In particular, it includes the distinction between receptive and productive vocabulary, implicit and explicit learning, item and system learning, explicit instruction, and lexical approach.

In second language vocabulary acquisition, there are two aspects of acquiring vocabulary: reception and production. According to Nation (2000), reception involves perceiving the form of a word in listening or reading while retrieving its meaning. On the other hand, production involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form. They are both important aspects in processing meaning for the reasons that if the learner aims to understand a text or a lecture, receptive knowledge of multi-word strings is essential. If the learner needs to use all the language skills, it is necessary to retrieve ready-made language patterns to maintain a discourse.

Apart from reception and production, vocabulary acquisition operates in two distinguishing factors in relation to psychological aspects (Ellis, 1997): implicit and

explicit learning. Implicit learning is defined as the acquisition of knowledge about the underlying structure of a complex stimulus environment by a process that takes place naturally, simply and without conscious operation. In contrast, explicit learning is a more conscious operation where the individual makes and tests hypotheses in search for a language structure. Hence, revealing the natural occurrence of formulaic sequences and continually bringing them to learners' attention would provide an extensive reach of vocabulary. Consequently, the more recognition of extensive patterns, learners might have good quality input and develop lexical nature of the language eventually produce longer patterns themselves.

In second language vocabulary acquisition, vocabulary knowledge can be divided into two schemes according to its linguistic aspects, item learning and system learning. According to Schmitt (2000), item knowledge is learning words as individual units while system learning is learning the systems or rules of language structures. Item learning focuses on the connection of words and phrases, for example, the utterance *by the way*, the phrase comprises three words when taken as a whole provides a single meaning. However, a disadvantage on item learning is that it is very gradual and requires a lot of exposure because learners have to memorize multiword units like individual words (Thornbury. 2002). On the other hand, system learning tends to analyze the detailed components and arrangement of the language for example, the verb 'talk' and its participle 'talk+ ed =*talked*.' Apparently, system learning concentrates mainly on the analysis of the segments and re-connecting the segments into chunks to produce a comprehensible language structure. Briefly, acquiring phonology, Cruttenden's (1981)

study showed that when children learn their mother tongue, they hear normal sequences, in this sense item learning is used. They tend to reduplicate clusters as unanalyzed units. Consequently, in advanced phase, they separate the sequence into fragments as analyzed units. After the units are identified, they are associated with another pattern to form a new sequence. Therefore, in comparison with second language learners, there is a possibility to acquire the language as multiword items then analyze them into individual items. Moreover, plentiful exposure of the language plus awareness of word combination patterns is direct approach focus of lexical approach and explicit instruction.

A recent extension of communicative language teaching is the lexical approach, an alternative for grammar-based approach that confers special attention to expressions and collocations. Traditionally, vocabulary and grammar are perceived as different entity while in lexical approach they are taken as a whole. As Lewis (1997) maintains that instead of words, language teachers consciously try to think of collocations, and to present these in expressions. Rather than trying to break things into ever smaller pieces, there is a conscious effort to see things in larger, more holistic, ways. He mentions further that the ability to chunk language successfully is central to understanding of how language works. It is due to the fact that native speakers retain many prefabricated lexical items in their memory, language learners fluency and accuracy could be achieved largely by retrieving and combining ready-made chunks of language. Furthermore, the approach focuses on the lexico-grammatical components of the language, taking words at the sentence level. An implementation of this teaching approach is largely directing learners in the observation on how multi-word expressions are arranged in the language. Leading

learners notice the occurrence of formulaic sequence, exposure, instruction and practice are also incorporated.

Nonetheless, explicit instruction, Thornbury and Slade (2000) describe three instructional model that allows language learners plentiful opportunities of exposure, and participation in authentic conversation. The following elements for the teaching of pragmatic competence include cognitive awareness, receptive skill development and productive use.

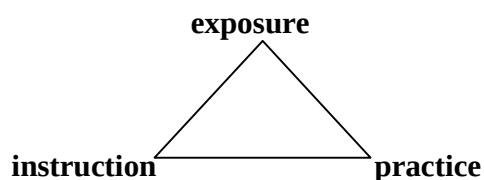


Figure 1 Elements of a pedagogy of conversation Adapted from Thornbury and Slade 2000: 295

The instructional model is arranged in a triangular shape because the elements assume a single sequence. Also, they can be organize in any order for example listening to audio recordings then followed by reading of a transcript. Alternatively, learners are given explicit instruction in the features of a conversation. They observe how these features work in context before going on to practice themselves. For instance giving instruction in the learning and memorization of formulaic sequences. The main feature is to raise learners' awareness of the formulaic sequences through frequent exposure and memorization (Lewis. 1997). In practice, learners would attempt a conversational task

then witness native speakers performing the task. Comparing learners' task between the native speakers, learners would notice the difference and are given instruction in these features before repeating the practice task. The challenge for a language teacher is to provide practice opportunities that target both accuracy and fluency. As Schmitt (2000), declares that explicit teaching in second language learning is essential for the most frequent words (formulaic sequences) because they are prerequisites for language use. Therefore, providing a lot of exposure on language features and directing learners to notice these features would give language learners with productive tasks.

Spoken Discourse Production

Additionally, the mental lexicon processes and organizes information in three complementary elements: input, storage, and retrieval (McCarthy. 1998). Firstly, input refers to linguistic input such as spoken data or stream of sound. The sources of input may be informal where vocabulary is acquired unconsciously. The learner, for instance, may learn words and phrases from media; entertainment and communication. Secondly, storage is a process in which words and phrases are organized to one another. Learned words or phrases are accumulated in multi-word patterns. Thirdly, retrieval is the ability to access to particular words and multi-word patterns when necessary without too much mental searching. Lastly, output is delivering the message. In the case, when speakers convey a message, they initiate with the information they want to evoke and have some patterns in the mental lexicon that can provide the major elements of this message. The patterns are stitched together with prefabricated multi-word strings to smoothen the

message then it is produced. Thus, both native speakers and learners mental processing exemplify formulaic sequences in their speech production focusing on a larger structure for fluency.

In terms of the number of words a speaker and listener need to control the demands placed on them are considerably few compared to writers and readers. Nation (1997) and Schmitt (2000) cite that a proficient speaker can largely function in everyday conversation with a vocabulary of 2000 words. According to Nation, a speaker can maximize such number of words in using repetitive words to achieve an acceptable degree of conversational competence.

In conclusion, the main advantage of formulaic sequences is to avoid processing overload for reason that teacher and learners alike could give close attention to formulaic sequences as basic existing unit necessary for comprehension and production. Another advantage in the teaching and learning of formulaic sequence is storage, learners tend to accumulate larger and wider groups of words. Considering the learning style in most second language Asian learners including Thailand, rote learning or learning by memory appears to be the common practice, a practice that could be an essential building block in the study of formulaic sequences. For instance instead of requiring learners to memorize individual words, a teacher could extend the words through their collocations, descriptive partners or contextual expressions. Taking into account intermediate learners state wherein they have learned or memorized individual words from this circumstance, a language teacher could instruct them to memorize multi-word patterns. Thus, increasing

their vocabulary in a sentence level through the memorization of repetitive multi-word patterns would broaden their production and use of the language.

Previous Related Studies

Formulaic sequences are a linguistic tool to any language learning and teaching. As a result, many research studies have been conducted in order to investigate their relevance in second language acquisition and production.

Tan (2002), observed the features of prepositional clusters into two distinctive aspects; meaning usage and grammatical distribution by using three different corpora to study the following clusters; *round about*, *in and out*, *ins and outs*, and their respective components; *round*, *about*, *in*, *out*, *ins*, and *outs*. She reported that there was a difference in the two aspects studied between the prepositional clusters and their corresponding components. For instance, the composite *round* takes numerous grammatical functions unlike the cluster *round about* which serves mainly an adverbial function. It was concluded that a vocabulary item perceived with a single meaning have particular meanings in other contextual situation for example *round about* as adverb modifier of place (we returned by the water side *round about the North-point*), *round about* as adverb modifier of time (It's going *round about 250 earth days*). An implication is that a single definition of a word is not sufficient to understand a meaning but rather consider the surroundings involved. It means that to learn and teach a new word, it should not be limited to a single word meaning but rather it is vital to include preceding words and the context.

Chamnan (2004) examined high frequency content words and their collocations in two civil engineering journals by using a corpus. She found that the collocations were two word combinations; mostly comprised of a variety of words such as adjective + noun, noun + noun, adjective + noun, and others. Furthermore, She also noted that every single word collocating with different words or phrases produces new meaning. Therefore, she expressed that collocations are a vital key in language learning. Teachers who actively introduce collocations are helping students be aware of natural co-occurrence of words in the language.

Comparing verb-noun English collocations between English and history textbooks, Koya (2004) analyzed the target collocations according to their kinds. She observed that the types of collocations in the history textbooks are slightly less than those in English textbooks. It was further stated that the difference in the contents was controlled by their contexts; concluding that there should be more variety of collocations included in textbooks to cultivate learners' better command of English.

Analyzing phrasal verbs in a newspaper's front-page headlines, Kittikool (2004) focused on idiomatic and literal meanings. He collected phrasal verbs in the newspaper's headlines, identified their conveyed meanings and gathered undergraduate students' feedback in understanding phrasal verbs. He found that both idiomatic and literal meanings were contained in the headlines. Frequency idiomatic meaning phrasal verbs appeared twice more than literal meaning. Therefore, the researcher concluded that meaning of phrasal verbs cannot be guessed by simply putting their individual units

together; a verb and a particle, rather have a handle on the verb combination as a single unit of meaning.

In summary, the studies reviewed reveal that the combination of words possesses meaning different from their individual component. In composition, FSs vary in the context while in frequency FSs display an array of functions. It indicates that the teaching and learning of formulaic sequences is fairly important to help learners in their speaking and listening, as well as reading and writing. In the meantime raising learner's awareness through exposure, instruction, and practice they would notice the diverse functions and usage of formulaic sequences. Therefore, in this study, the researcher mainly focuses on identifying types of FSs, and their frequency in conversation exercises; hypothetically, identifying FSs and their frequency would be of relevance in vocabulary teaching and learning.

CHAPTER 3

METHODOLOGY

This chapter provides an overview of this study in four main sections: data collection, research instruments, procedure, and statistical treatment.

Data collection

The data were selected transcripts of the three English university textbooks entitled: *English Upgrade 2* (Lambert, Gershon & Mares. 2002), *Interchange 2* (Richards, Hull & Proctor. 2005) and *Person to Person 2* (Richards, Bycina & Wisniewska. 2005). The three English textbooks were chosen according to the following criteria. Firstly, they are designed commonly for university learners specifically ranging from beginners to intermediate levels (wherein the target level of this study is basically the intermediate level textbooks since the material writers labeled them as *Book 2*). Secondly, the unit contents focus on different language functions and structures covering the necessary language patterns and vocabulary exemplifying a sample of mainstream in general conversational English. Lastly, they contain specific exercises and activities on the four language skills, basically, allowing learners to integrate the language learned in their practice. In addition, the three English university textbooks are published by different publishers namely: *English Upgrade 2* by Macmillan Publishers, *Interchange 2* by Cambridge University Press, and *Person to Person 2* by Oxford University Press.

The data were collected from the conversation sections excluding pronunciation exercises and answer keys to avoid overstatement on the results. The total number of

transcripts were recorded as follows; 24 in *New Person to Person*, 26 in *English Upgrade 2*, and 31 in *New Interchange 2*, 91 transcripts in total. The transcripts include wide range of conversation topics from formal to casual pertaining on the following contexts; social interactions, locations and directions, requests and permission, suggestions and advice, occupations, hobbies, news and entertainment, culture, festivals and holidays, and science and technology. The conversation practice also includes various communicative skills such as describing someone or something, expressing opinions, suggesting, narrating, and negotiating ideas. It also includes communicative strategies such as confirmation, affirmation, repetition and questioning.

Research Instruments

Research instruments describe the concordance program and the eight categories of FS used.

Concordance Program

The *Simple Concordance Program 4.0* (Reed, 2006) is a windows program for linguistic research that generates in computer-based files. Aside from handling a large number of texts, there are three specific tools utilized in this program. First, it can understand sequences of words in many different ways. Second, it can sort words according to their frequency and percentage of occurrence in a certain context. The following is an example of the word *as* from the British National Corpus (BNC) written text.

Cluster of as	Frequency	Percentage (100%)
as * as the	4566	53
as well as the	2674	31
as far as the	874	10
as long as the	316	3.7
as much as the	50	0.6

Figure 2 Frequency and Percentage of the Clusters of *as*

The first column shows the sequences that follow *as*. The second column gives the total frequency of all the phrases, and the third column indicates the percentage of the phrases in the entire corpus.

Another tool of a concordance program is that it can identify clusters of words from a single word displaying the corresponding co-occurrences of a target word to the required number of collocates as in the sample of the word *to* the second highest frequency word in *Interchange 2*.

```

89      It can sometimes take hours to go from one part of the
90      go from one part of the city to another./When I don't
91      last train home or don't want to go all the way home of
92      to go all the way home only to turn around and come
93      to turn around and come back to work again. So, it gets
94      you recommend a capsule hotel to other people?/BRAD: Sur
95      are small, but you get used to sleeping in a small s.
96      recommend a capsule hotel to people who can't relax
97      Listen again. In addition to a bed, what else does
98      What changes would Brian like to make in his life?/BRIAN
99      It's sometimes pretty hard to pay the rent. I'm think
100     kind of job would you like to look for?/BRIAN: I'm
101     sick of this place. I need to live somewhere more exc
102     Hey, maybe we could move to another city. We could
103     me. Urn, I think I'm going to order something else, i
104     . I'm waiting for the waitress to bring me a fork./
105     Listen to people explain how to make these snacks. W
106     from 1 to 4./1./WOMAN: I love to eat this at the movies.

```

Figure 3 Concordance lines of the preposition *to*.

The concordance indicates citations of *to* as shown in *Figure 3*. In this case the concordance underpins the analysis of the instances of the structures that are connected to the target word.

Categories of Formulaic Sequences

Formulaic sequences (FSs) are categorized into two major types: polywords and multi-word expressions. They are further subdivided into eight sub-types. The classification of the categories is a combination of Lewis's (1997:7-11) categories on lexical phrases, and Moon's (1997) types of multi-word units. This is due to the fact that combining both classifications encompasses the widespread formulaic sequences in consideration of their semantic, syntactic and pragmatic characteristics.

A. Polywords are permanently combined two or three words most likely carrying a single meaning.

1. Binomials- fixed pairs of word connected by the conjunction, 'and'.

Examples: *salt and pepper*

black and white

2. Compounds – words which combines two or sometimes more different words. Frequently, the meaning of a compound word is conveyed in the two combined words. Compounds can be written either as one word or as separate words.

Examples: *bodyguard*

floppy disk

3. Collocations – words or phrases which is frequently used with another word

or phrase, in a way that sounds correct to native speakers but not be expected from the meaning.

Examples: *a hard frost*, 'hard' is a collocation of frost but not 'strong'

a heavy rain, 'heavy' is a collocation of rain but not 'light'

4. Phrasal verbs- phrases consisting of a verb in combination with a preposition or adverb or both, the meaning of which is different from the meaning of its separate parts.

Examples: *look after* means to take care of someone or something

work out means to exercise

B. Multi-word expressions are sets of phrases and sentences that have fixed or semi-fixed word combinations.

5. Prepositional expressions – groups of words consisting of prepositions and locations or places.

Examples: *in the corner*

top of the table

6. Fixed expressions – words or groups of words used in a particular situation or by a particular people.

Examples: *I think so*

For fun

7. Semi- fixed expressions - phrases or sentences with one or more variable slots which can be filled by an item chosen from a relatively small groups of items which share particular language characteristics.

Examples: *It's _____ than I thought.*

Could you pass the _____ please?

8. Idioms - groups of words in a fixed order that have a particular meaning that is different from the meaning of each word understood on its own.

Examples: *on top of the world* means 'feeling pleased or happy'

lend someone a hand means 'to help someone'

Procedure

This study is a quantitative type focusing on the analysis of the occurrence of formulaic sequences. This section embarks on analysis of the collected data from the three English university textbooks.

Data Analysis

The analysis of the data was performed in two ways in order to answer the two research questions in this study.

First, to identify the categories of the formulaic sequences found in the target English university textbooks, the following steps were observed.

1. The transcripts in each textbook were examined manually in order to identify the formulaic sequence (FSs).
2. Each identified FS was sorted into their corresponding categories.
3. The eight categories were presented in tables.

Second, to obtain the frequency of formulaic sequences in the three English university textbooks, the frequency of the FSs were analyzed using the *Simple Concordance Program 4.0*. The following steps were carried out.

1. All ninety-one transcripts were scanned and converted into electronic files.
2. Pronunciation activities and answers for listening exercises were excluded in order to avoid affectation on the frequency of FSs. For example the reference title: *Tape script 31 Conversation 1* occurs most often.
3. The files were inserted into the source file section of the program one at a time.
4. Frequency sort was done first in order to get the ten frequent words in the three university textbooks.
5. Each of the top ten frequent words was used as nodes for analysis.
6. The top ten frequent words were keyed in the program one at a time with the Phrase Command set to 5-word category.
7. The range of phrases was limited in three occurrences.

Statistical Treatment

To determine the significance in the occurrence of formulaic sequences in the English university textbooks, the frequency of formulaic sequences was treated with the percentage. The overall total of the formulaic sequences from the three English University textbooks was calculated with a full hundred percent.

$$\frac{\text{Total number of FSs in each category}}{\text{Total of FSs in all three textbook}} \times 100 \%$$

Total of FSs in all three textbook

Furthermore, the results were presented in tabular form for clarity and ease of access.

CHAPTER 4

FINDINGS

This chapter presents the results of the analysis of the occurrence of formulaic sequences in the three English university textbooks and in conformity to answer the following research questions:

1. What are the formulaic sequences found in the transcripts of the target English university textbooks?
2. How frequently do the formulaic sequences occur in the target English university textbooks?

The following tables show the percentage on the occurrence of frequently used formulaic sequences found in each English language university textbook. A figure is added to display the occurrence in the three English language university textbooks. Also, it includes the list of the top ten frequent words found in the concordance program followed by samples of formulaic sequences and their corresponding concordance lines occurring in the corpus.

Table 1 Formulaic sequences occurrence in the three English language
university textbooks

Category of FSs	Frequency of Occurrence	Percentage (100%)
Polywords	137	24.20
Multi-word expressions	428	75.80
Total	565	100%

In Table 1, it presents the total number of formulaic sequences in the three English university textbooks. It can be seen that multi-word expressions had highly more occurrences with 79.80% in comparison to polywords with 24.20% occurrences. An indication is that multi-word expressions build up most of the patterns in a spoken discourse while polywords austere provide accurate information to complete a speech. For instance in exchanges wherein listeners respond to confirm or agree on a prior statement, most responses are in a sentence level such as *That's a great idea! You've talked me into that.* Nonetheless, the difference of their occurrences, multi-word expressions and polywords have complementary functions because in spoken language, speakers tend to use them creatively and comprehensibly in order to relay a clear and accurate message to listeners.

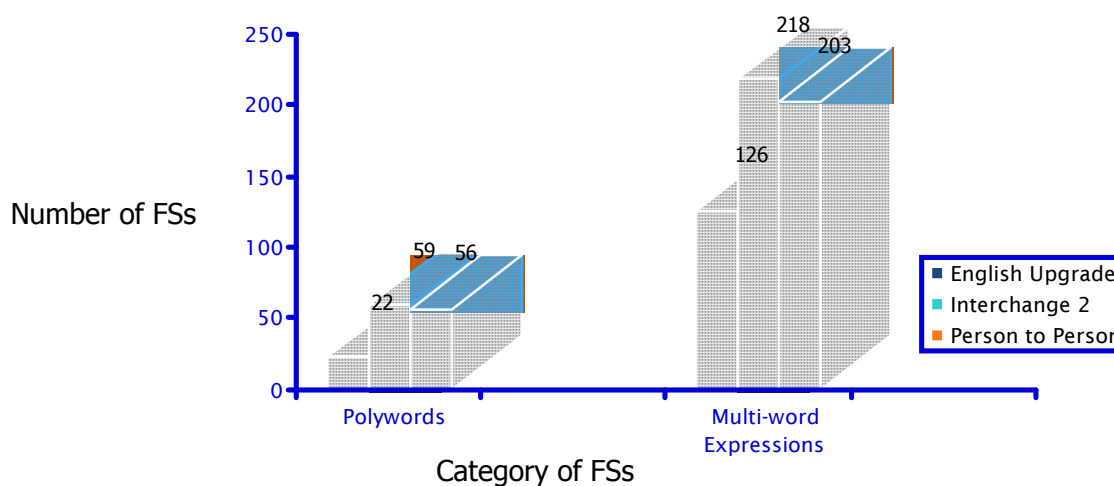


Figure 4 Total number of formulaic sequences in the three English university textbooks

In Figure 4, it illustrates the number of formulaic sequences in each English university textbooks. Perceptibly, Interchange 2 had the most number of formulaic sequences while English Upgrade 2 had the least number. The differences in the numbers of formulaic sequences display how each material developer applies the language differently. It seemed that each developer has their own choice of formulaic sequences and diverse presentation styles in each conversation exercise.

Table 2 Frequency of polywords occurrence in the three English language university textbooks

Category of polywords	Frequency of occurrence	Percentage (100%)
1. Phrasal verbs	45	32.85
2. Compounds	39	28.47
3. Binomials	29	21.17
4. Collocations	24	17.52
Total	137	100 %

In Table 2, among the four categories of polywords, it can be seen that phrasal verbs had the most occurrence (32.85%) followed by compounds (28.47%), binomials (21.17 %), and collocations (17.54%) See *Appendix C* for samples. Apparently, the occurrence in polywords had a fair distribution among the four categories for the reason that their disparity were within ten percent difference, 4 % and 7 % respectively. An indication is that the use of polywords as formulaic sequence is relevant in spoken discourse for they are systematically chosen to complete a particular context.

Table 3 Frequency of multi-word expressions occurrence in the three English language university textbooks

Category of multi-word expressions	Frequency of occurrence	Percentage (100%)
1. Fixed expressions	272	63.40
2. Semi-fixed expressions	111	25.87
3. Prepositional phrases	43	10.02
4. Idioms	2	0.82
Total	428	100 %

In Table 3, among the categories of multi-word expressions, it can be seen that the occurrences of multi-word expressions; fixed expressions had the most occurrence (63.40%) followed by semi-fixed expressions (25.87 %), prepositional phrases (10.02%), and idioms (0.82%) had the least occurrence. Apparently, the disparity of the occurrence among the four categories had enormous differences; 10 %, 15%, and 38% respectively. An indication is that native speakers use more fixed expressions more than the other three categories in their spoken discourse for instance in opening and closing a conversation, turn-taking or exchanges.

Table 4 Ten high-frequency words from the English language university textbooks
(2471 Words)

Word	Word frequency	Number of Words	Percentage (100%)
a	696	1	3.6
the	594	1	3.1
I	542	1	2.8
you	526	1	2.7
to	497	1	2.7
and	407	1	2.6
in	239	1	1.2
it	220	1	1.1
is	215	1	1.1
of	184	1	1.0

In Table 4, the ten high frequency words are presented with the number of frequency and their percentage out of 2471 words in the entire spoken corpus. The two highest frequency words are the articles *a* and *the* followed by the pronouns *I* and *you* the preposition *to* then the conjunction *and*, and then the helping verb *is*. Seemingly, all the listed most frequency words are function words; articles, pronouns, prepositions and helping verb. The results confirm Stubbs's (1996) that in a spoken language, more function words are used more than content words making the discourse more predictable. Primarily, function words can be freely inserted into a conversation in which they serve an interactive and interpersonal purpose. In sum, function words are commonly used in spoken language for the reason that they fill the informational gaps in a speech.

The following tables are sample concordance lines of the eight sub categories of formulaic sequences as they occur in the corpus.

Table 5 Sample concordance lines of collocation

Key word	Sample concordance lines
take	
0884	for anything. We should take a bottle of aspirin . We could
0886	DANA: OK. Should we take anything for cuts? I always
0888	when I camp. /HILARY: Let's take a tube of ointment. DANA
1000	for anything. We should take a bottle of aspirin . We could
1002	DANA: OK. Should we take anything for cuts? I always
1004	when I camp. HILARY: Let's take a tube of ointment. DANA
1010	a lot. HILARY: We should take this lotion. I've used it
1012	works. DANA: Should we take sunscreen? I don't want to
1269	, where you bought it. They'll take care of it for you. M: Back
1281	I'll get her. 2. Offering to take a message A: Hello. B: Hi
1284	not here right now. Could I take a message ? B: No, thanks
1289	not here right now. Can I take a message ? B: Yes, please
1298	isn't home right now. Could I take a message ? P: This is Pete
1308	She's not in right now. Can I take a message ? E: Yes, please
1360	a customer right now. Can I take a message and have her call
1427	overpackaging! We should also take our own bags to the grocery
1563	headache. A: You'd better take some aspirin . B: That's a
1573	go out today. Why don't you take some aspirin and lie down
1576	T: I'm not sure. I'll take a look. Or why don't you use
1614	food. A: Am I allowed to take aspirin with this medication
1616	B: No, you shouldn't take any aspirin .

In Table 5, collocates of *take* for *take a message* and *take aspirin* appeared more than once in the corpus. *Take aspirin*, is interesting because it occurs in three different structures, *take some aspirin*, *take any aspirin* and *take a bottle of aspirin* while *take a message* remained the same in Lines 1281, 1284, 1289, 1298, 1308 and 1360. An indication is that the occurrence developed in the same order of lines in the corpus. Apparently, functions of the mentioned collocations on the word *take* were mostly giving suggestion and medical advices in relation to health problems.

Table 6 Sample concordance lines of binomials

Key word	Sample concordance lines
<i>and</i>	0038 code ... Clerk: Good evening and welcome to the eight-one-eight 0079 ? This guy with <i>the ponytail and the beard</i>? Tina: Oh, that's 0109 three times a week. mustache and big glasses? Julie: Aikido 0179 I don't have to wear a <i>suit and tie</i> to work. Unit 5 Having 0638 ANNE: Well, how about <i>jeans and a cotton shirt</i>? JOSH: No, I 0639 , what about your <i>black pants and silk shirt</i>? /JOSH: That 0647 blue one is nice. JOSH: OK, and I'll bring my leather jacket 0651 invite friends to events and activities. Complete the chart 0784 . She's wearing <i>black pants and a white shirt</i>. She's very classic 0868 to walk around. But it's <i>hot and humid</i> there. MAYA: Hmm. I 1318 fill out an application form and send it to us with the registration . 1417 , <i>ladies and gentlemen</i>, and welcome. I'm sure we have all 1420 families buy paper <u>napkins and towels</u> at the supermarket 1424 can start using <u>cloth napkins and cloth towels</u> instead of paper 2028 spent /hours running <i>up and down</i> the beach with it first, and

In Table 6, the binomials *pants* and *shirt*, *napkins* and *towels* occurred at least twice in the corpus. Interestingly, *pants* and *shirt* has a variation in Line 0638, *jeans* and cotton *shirt* understandably, *jeans* refers to *pants*. Although the sequence has a variation the order of the elements are still irreversible. As Carter and McCarthy (1988) stated that neither way is more logical than the other, they are probably seen simply as fixed. Moreover, the repetitive use of these elements in a discourse makes the pattern arbitrary.

Table 7 Sample concordance lines of Compounds

Key word	Sample concordance lines
<i>sun</i>	
	0417 Dr. Health: Right. You have sunburn . You should rub some aftersun
	0419 should always use a strong sunblock cream when you're at
	0421 the worst time, when the sun is the strongest. Caller
	0896 works. DANA: Should we take sunscreen? I don't want to get
	0897 ? I don't want to get a sunburn . HILARY: Definitely.
	1012 works. DANA: Should we take sunscreen? I don't want to /get
	1013 ? I don't want to get a sunburn . HILARY: Definitely.
	1580 really work. P: Too much sun is dangerous, you know. M:
	1583 gel...if we have P: Try this sunburn lotion . It'll help 2 any
	1645 need something for this sunburn .
	1735 don't wear a hat, you'll get sunburned .
	1739 should have a pair of strong sunglasses to protect your eyes
	1743 to wear them on really warm sunny days. And how could I forget
	1746 bring some so you don't get a sunburn . Are you bringing a nice
	2174 dark. There was a beautiful sunset. I'd had a very difficult

In Table 7, the compounds for *sun* had been noted such as *sunburn*, *sunblock*, *sunset*, *sunglasses*, and *sunscreen*. In these samples of compounds, the head word is *sun* combined with modifiers *burn*, *block*, *set*, *glasses* and *screen*. Having fixed form and syntactic regularity, compounds are considered prefabricated set of words that are naturally and easily accessible in speakers' discourse.

Table 8 Sample concordance lines of Phrasal verbs

Key word	Sample concordance lines
up	0063 , don't worry. We'll shape you up in no time. Unit 2 Family and 0135 day? You know, when you get up, when you go to work, stuff 0136 Well, let me see ... I get up really late - around noon. 0222 Sounds good. I'll pick you up at university? seven at your 0239 dad's German. But I grew up in New too. When are you thinking 0366 few years, but we didn't give up. And now, well, our business 1396 this is Ron. Something's come up and I can't make it to the show 1527 hear about Brenda? She ended up at the hospital. She was... 1540 What did they do? They put up their tent to try and keep warm 1689 that. S: My knee has swollen up and it's really bad when I go 1842 from them. They'll send it up to your room. G: OK, thanks 1900 here where we picked you up, outside the Walker Art Center 1903 well. But you may well come up with a question I can't answer 1911 thing you have to do is blow up the balloon. Next, cover the 2029 I got the lines tangled up right away. Once you're up on 2121 A: I agree. Kids grow up thinking that it's OK to hurt 2176 evening at home. I looked up at the sky, and I saw this string

In Table 8, the phrasal verbs with the preposition *up*, obviously most of the sequences appeared once except for *come up*, *grow up*, *pick up* with two occurrence while *get up* for three times. Seemingly, the composition of phrasal verbs varies from separable to inseparable features for instance Lines 0135 and 0136, *get up* comes in random order while Lines 0222 and 1900, the phrase *pick up* is separated with you.

Table 9 Sample concordance lines of Prepositional Phrases

Key word	Sample concordance lines
<i>in the summer</i>	0494 Do you have any plans for the summer? Rachel: Well, I'm planning 0533 mall? But there is more! Next summer a fantastic vacation in 0567 it over there. In fact last summer, I went to Spain with my 0828 but I lived in Chile for a summer when I was in high school 0937 And how hot is it in the summer? KAr: Well, up in the mountains 1055 And how hot is it in the summer? KAr: Well, up in the mountains 1057 NINA: What about rain? KAr: Summer is the rainy season, and 1088 and we moved to New York last summer. WILL: That's great. Are 1394 Bye. 2V: Hi. This is Dave Summers. I can't talk right now 2173 about four years ago in the summer. I was driving home from

In Table 9, the key word for prepositional phrase is *in the summer*, interestingly the target sequence appeared three times in Lines 0937, 1055 and 2173. It is also observed that the sequence has other variations such as *for the summer*, *for a summer*,

next summer and *last summer*. Functionally, the variations in the prepositional phrases referred to time the speech happened for instance *for a summer* and *last summer* both phrases are directed in the past while *for the summer* and *next summer* directed in the future.

Table 10 Sample concordance lines of semi-fixed expressions

Key word	Sample concordance lines
<i>that sounds</i> _____	
	0110 that sounds pretty Tina: Oh, that's my Uncle Sam, interesting
	0128 a week. Julie: Aikido, hey that sounds pretty interesting
	0142 keep in shape. Interviewer: That's a kind of martial art, right
	0216 , I really need some Pablo: Oh that's too bad. Well, how about
	0243 : Saturday morning? Yeah, that's good. Lenore: Ah, so you're
	0258 doing? Lenore: I'd love to. That would be excellent. Tony
	0380 or cologne? Maria: Hey, that's not a bad idea. Ann: Yep
	0496 paint my room ... Tom: Hmm, that doesn't sound very exciting
	0502 do a lot of sailing. Tom: That sounds fabulous! Rachel:
	0507 Rachel: A cruise! Wow, that sounds great. Tom: I know, but
	0641 a silk shirt? JOSH: That sounds better. But it's cold tonight
	0916 NANCY: So do I. Hmm. That sounds good. I'll think I'll have
	2068 there, terrified. P: That sounds awful. How embarrassing

In Table 10, the semi-fixed expression *that sounds* _____ is filled with different words on emotions. The expression is substituted according to the context of the conversation for example in Lines 0128, *that sounds pretty interesting*, substituted in

Line 0502 and 0515 as *that sounds fabulous*, while in Lines 0916, 0641 and 2068 the expressions are substituted with *better*, *good* and *awful*. Logically, in an interactive conversation, speakers and listeners confirm, affirm, compliment or comment a particular sequence to maintain a communicative activity.

Table 11 Sample concordance lines of fixed expressions

Key words	Sample concordance lines
<i>could you tell me</i>	0022 Staff member: OK, first could you give me your Kyoko: 0046 Clerk: Fine. Could you please fill in the four-nine-oh 0047 registration form? And could you tell me your e-mail address? 0172 THREE Interviewer: Eric, could you tell me about your present 1239 directions in a store (1) A: Could you tell me where I can find 1283 a message A: Hello. B: Hi, could I please speak to John? 1284 he's not here right now. Could I take a message? B: No 1298 . He isn't home right now. Could I take a message? P: This 1299 competition next weekend. Could you ask him to give me a 1317 in taking a language class. Could you tell me how to apply 1320 registration fee. B: Great. Could you send me a form, please 1812 A: Can I help you? B: Could you tell me what time the 1820 Here we are. For five nights. Could you fill out the registration 1936 midnight. W: I wonder if I could get a wake-up call for 6

In Table 11, the expression *could you tell me* appeared at least five times in the corpus. Obviously, the function of the expression is to ask for request therefore, the

occurrence in Lines 0047, 0172, 1239, 1317, and 1812 represented different contexts such as asking for personal information, directions, and local time. The recognition of the recurrent feature of fixed expression asserts that they occur as a formulae and repetitive in a particular situation. Additionally, in second language teaching, for learners to adapt these expressions it's better explained in their use rather than in their semantics. They need to explore wide variety of communicative expressions to meet a wide variety of communicative demands.

CHAPTER 5

CONCLUSION AND DISCUSSION

In this chapter, the data analyzed are discussed to answer the research questions. It includes conclusion and discussion as well as pedagogic implications, limitations in the study and suggestions for further studies.

Conclusion

This study aimed to examine the feasibility of identifying formulaic sequences worth teaching to second language learners. It analyzed those sequences that occur in the transcripts of three English university textbooks. The results revealed that a total number of 565 formulaic sequences in which 426 (72.8%) are multi- word expressions and 137 (27.2%) are polywords. Classifying formulaic sequences provided a perspective on how native speakers use prefabricated words as sequences and their corresponding functions. Consequently, in teaching vocabulary and conversation, language teachers would focus on different priorities and emphases on formulaic sequences.

Discussion

This study aimed to answer the following research questions.

1. What are the formulaic sequences found in the transcripts of the target English university textbooks?

The results showed that out of the 565 formulaic sequences (FSs) found, multi-word expressions was relatively high in occurrence than polywords. Among the

eight sub-categories, fixed expressions had the highest occurrence (272) followed by semi-fixed expressions (111) and idioms had the lowest (2) in which interestingly they belong to multi-word expressions category.

The results imply that the pervasive occurrence of fixed and semi-fixed expressions underpin the view that for speakers to convey a message, they facilitate their conversation with prefabricated multi-word patterns (Pawley & Syder. 1983). It has been observed that fixed expressions are used to maintain the conversation while semi-fixed expressions are composed of interrogative sentences wherein the slots can be filled with vocabulary needed for the situation. Evidently, the observation supports Sinclair's (1991) distinction of multiword expressions between open choice principle and the idiom principle. Open choice principle is when speakers use the language creatively through filling in slots. This principle corresponds to the function of semi-fixed wherein speakers can add or substitute information necessary for their speech. For instance the semi-fixed expression *the best thing*, *thing* is substituted with *part*. Idiom principle on the contrary, is when the pattern could not be substituted; changing the order of the sequence or some of the elements would convey a different meaning. In this study, fixed expressions and idiom fall into this principle, for example the fixed expression *don't mention it*, and the idiom *the more the merrier*, an alteration of the element like *the less the merrier*. In sum, Nattinger (1988) mentions that the retrieval of formulaic sequences permits a direct attention to the larger structure of the discourse rather than focusing narrowly on individual words. Moreover, formulaic sequences are used in the speech functions through interactional and lexical relation. Interactional, specifically for fixed and semi-fixed expressions,

they function in a sense that a conversation is maintained by performing as short cut processing devices and frameworks (Wray. 2000).

On the lexical relation, a conversation is reinforced with the use of binomials, collocations, compounds, phrasal verbs, prepositional phrases and idioms including simply individual words. An implication is that a conversation is carried on with fixed expressions and semi fixed expressions as frameworks, then integrated with polywords, prepositional phrases, idioms and individual words in relevance to the context. Consequently, the lexical relation differs within contexts as proficient speakers have their own choice of words to utilize in a certain conversational topic. Similarly, McCarthy & Carter (1997) describe the characteristic of spoken language that it uses a lot of interpersonal phrases, organizational markers, and other kinds of discourse items. By its very nature and constraints, spontaneous spoken language involves heavy use of formulaic sequences. In conclusion, the occurrence of formulaic sequences in the conversation transcripts of the three English university textbooks, essentially, both multi-word expressions and polywords operated incorporate to convey a comprehensible message in both discoursal and lexical aspects.

2. How frequently do the formulaic sequences occur in the target English university textbooks?

To confirm the occurrence of the formulaic sequences in the English university transcripts, a spoken corpus was compiled. With the use of concordance program, ten high frequency words listed, most of which are function words that are regularly present in the formulaic sequences. Articles, prepositions, pronouns,

conjunctions, and forms of verb *be* are equally necessary regardless of the topic. In particular, the ten high frequency words did not occur as individual words but turn out to be elements of formulaic sequences. In view of article *the*, the second top high frequency in the word list, it is used before a new content word.

1214 a new shirt? B: **You can try the store** on Washington Street
 1214 Street. It's next to the King Building. 2. Describing
 1216 buildings A: **Which one is the King Building?** B: It's the
 1217 the King Building? B: It's the big glass office building just
 1217 office building **just past the park.** LISTEN TO THIS PAGE
 1219 12 B: Hi, Kumiko. Hey! **What's the matter?** K: It's my boyfriend's
 1222 glove. I just went to the Athletic Center, but I couldn't

In general and spoken English, function words tend to be frequently used although they hold no or little meaning, they contribute in the grammatical structure of the language (Schmitt. 2000). Therefore, the more language learners are acquainted to the frequently occurring phrases and words the fewer types need to be known to understand a conversational topic.

Another observation in this study is that, it was expected that the frequency of a formulaic sequence would range to at least four times but surprisingly there was a diverse frequency collected. The occurrence of formulaic sequences was varied in each category for instance with phrasal verbs *get up* and *pick up* (see Table 8), they occurred twice in the corpus while the expression *could you tell me* occurred more than four times. The difference on the various distribution of occurrence in the formulaic sequences merely indicates how material developers thoroughly design and select the content of the transcripts. They have their own lexical choice as well as expression that could be situational or culturally related in the contexts.

Pedagogic Implications

In second language classroom for language learners to communicate competently and efficiently, their discourse should be supported with formulaic sequence knowledge. According to Thornbury (2006), multiword expressions and polywords may provide a learner with a more effective bridgehead into the spoken language. It is essential, therefore, to reconsider the approach in teaching vocabulary and spoken discourse. In reference to teaching, teachers and material developers can increase learners' formulaic sequence knowledge through raising consciousness. Raising learner's awareness can be implemented in a combination of instruction, exposure and authentic practice. Practically, transcripts are valuable in instruction for a reason that they are the most likely available for a focus on the specific features of spoken language. Willis and Willis (1996) point that using transcripts allow learners time to notice features that could not be heard in real time conversation. Some classroom tasks associated with the use of transcripts include: underlining or counting instances of formulaic sequences, gap filling as a post- listening task, and reading aloud. A more explicit instruction might usually focus on learning and memorizing formulaic sequences. To mobilized fluency, instructions should involve learning and controlled practice of multiword expressions and word combinations. Thus, these techniques focus on specific features of the spoken text, they constitute both consciousness raising and form of instruction.

Sensibly, in the practice stage some of the teaching strategies could be facilitated, for previously learned sequences ask learners to verbally rehearse them. Recycling would help learners store the sequences (Nation. 2000). Another teaching strategy is sentence reading; here, learners are shown with some samples of the pair of

words in simple sentences. They are told to read the sentences, and later see how the pair of words is associated. Sentence generation learners are shown each pair of words and asked to construct and say aloud a meaningful sentence relating to the word combinations (Thornbury. 2006). Seemingly, one logical teaching technique is exposure, in the form of listening to audio recordings accompanied or followed by the reading of the transcript. An extension of this activity is that learners could try to reconstruct the transcript themselves. Exploiting learners in listening audio recordings or video extracts could be an alternative for a native speaker absence in the second language classroom. In conclusion, the teaching strategies presented on exposing learners to the formulaic sequences through the use of transcripts, either in listening or reading practices could be adapted in other second language classrooms. However, these strategies are not ends to the means, teaching vocabulary in conversation could be incorporated with the use of corpora.

The use of corpus evidence in second language teaching supports learning for the reason that it can inform language teachers better on their presentation about language structure and their components, they can also provide clearer and more authentic samples of this structure. For example in using the concordance the teacher can understand the subtle differences between the forms and usage of a sequence. Since the rise of corpus analysis, second language teaching has been considerably affected results from investigations in both native speaker's and learner's corpora were used to design materials and framework for classroom instructions. Consequently, for English language teachers, one aim in teaching is to increase the learner's confidence, therefore in vocabulary teaching is it should not be focused mainly on acquiring new vocabulary in speaking, but teaching learners how to make

sentences making use of the input acquired through listening and reading. In practicing the language both controlled and free procedures should be included wherein learners can produce the acquired chunks (Lewis. 2000). Finally, for learners to produce chunks rather than focusing on grammar they should explore various communicative modes with patterns in a more generalize sense and adapt them in their speech.

Limitations of the Study

The limitations of this study are enumerated as follows:

1. This study is limited on the proportion of the corpus. Ideally, a corpus must be balance in including different genres of a language; the data collected was concentrated in conversation exercises specifically dialogues.
2. This study is controlled in the number of textbooks. There are only three English language university textbooks as it could be a restriction in the representation of formulaic sequences used in spoken discourse.

Suggestions for Further Research

This section provides some suggestions on how to supplement this study.

1. There should be a study on the learners' actual use of formulaic sequences to test if the formulaic sequences in the intermediate level are attainable.
2. There should be a parallel comparison on native speakers' spoken corpora against non-native speakers. In order to disambiguate structures which are used for discoursal and lexical components.
3. In vocabulary learning and teaching, data driven learning using corpora

should be incorporated, basically, using a concordance program because it could precisely define confusing words and clearly identify the component words in a particular context.

4. There should be further studies on other levels of English university textbooks focusing on the speaking section, as well as focusing on the other language skills.

5. There should be further studies on other genres of spoken discourse such as report, debate, oral presentation, etc.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Allwright, R.L. (1981) What do we want teaching materials for? *ELT Journal* 36 (1):5–18.
- Baigent, M. (1999) Teaching in chunks: integrating a lexical approach. *Modern English Teacher* 8 (2): 51–4.
- Biber, D., Conrad, S. & Cortes, S. (2004). If you look at: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25 (3): 371-405.
- Biber, D., Conrad, S. & Leech, G. (2002). *Longman of Spoken and Written English*. Harlow: Longman.
- Biber, D., Conrad, S. & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics* 15 (2): 169–89.
- Bishop, H. (2004). Noticing Formulaic Sequences- A Problem of Measuring Subjective. *Linguistics* 4:15-29. Retrieved February 2, 2005, from <http://ling.wisc.edu/lso.wpl/4.1/LSOW4.1-04-Bishop.pdf>.
- Bycina, D., Richards, J. & Wisniewska, I. (2005). *Person to Person*. Student Book 2. Hongkong: OUP.
- Bygate, M. (1987). *Speaking*. Oxford: OUP.
- Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal* 52 (1): 43–56.
- Carter, R. & McCarthy, M. (1995). Grammar and the Spoken Language. *Applied Linguistics* 16 (2): 141-157.
- (1998). Developments in the Teaching of Vocabulary. In R. Carter & M. McCarthy (eds.). *Vocabulary and Language Teaching*. 39-51. London: Longman.
- (1998). *Vocabulary and Language Teaching*. London: Longman.
- Celce-Murcia, M. & Ohlshtain, C. (2000). *Discourse and Context in Language Teaching: A Guide for Language Teachers*. Cambridge: CUP.
- Chamnan, P. (2004). *A Corpus Study of High-frequency Words in Civil Engineering Research Articles: Subdisciplinary Differences between Structure and*

- Transportation*. Master's Thesis. M.A. (Applied Linguistics). Nakhorn Pratom. Graduate School Mahidol University. Photocopied.
- Cowie, A. P. (1998). Stable and Creative Aspects of Vocabulary Use. In R. Carter. & M. McCarthy (eds.). *Vocabulary and Language Teaching*. 126-139. London: Longman.
- Cobb, T. (2000). *The Compleat Lexical Tutor for Data-Driven Learning on the Web*. Retrieved November 2, 2004, from <http://132.208.224.131>.
- Cruttendens, A. (1981). Item-learning and system-learning. *Journal of Psycholinguistic Research*. 10 (1) :78-88.
- Ellis, N. (1997). Vocabulary acquisition: word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge. CUP. 23-50.
- Fernando, C. (1996). *Idioms and Idiomacity*. Hongkong: OUP.
- Fletcher, W. (2004) *Exploring Words and Phrases from the British National Corpus*. Retrieved February 2, 2004, from <http://pie.usna.edu>.
- Flowerdew, L. (2003). A Combined Corpus and Systemic-Functional Analysis of the Problem-Solution Pattern in a Student and Professional Corpus of Technical Writing. *TESOL Quarterly* 37 (3):489-509.
- Frazier, S. (2003). A Corpus Analysis of Would-Clauses Without Adjacent If-Clauses. *TESOL Quarterly* 37 (3): 443-465.
- Fox, G. (1998) Using corpus data in the classroom. In B. Tomlinson (ed.). *Materials Development in Language Teaching*. Cambridge: CUP. 25-43.
- Islam, C. & Timmis, I. Lexical Approach 1-What does the lexical approach look like? *Teaching English*. Retrieved February 5, 2005, from http://www.teachingenglish.org.uk/think/methodology/lexical_approach1.shtml.
- Gershon, S. & Mares, C. (2002). *English Upgrade*. Student Book 2. Oxford: McMillan.
- Gilmore, A. (2004). A Comparison of Textbooks and Authentic Interactions. *ELT Journal* 58 (4):363 -374.
- Hull, J., Richards, J. & Proctor, S. (2005). *Interchange*. Student Book 2. Hongkong. CUP
- Howarth, P. A. (1996). *Phraseology in English academic writing: some implications for*

- language learning and dictionary-making*. Tübingen: Niemeyer.
- (1998). Phraseology and Second Language Proficiency. *Applied Linguistics* 19 (1): 24-44.
- Hunt, A. & Beglar, D. (2002). Current Research and Practice in Teaching Vocabulary. In J. Richards & W. Renandya. (eds.). *Methodology in Language Teaching: An Anthology of Current Practice*. Cambridge: CUP. 16-34.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: CUP.
- Huttner, J. (2005). Formulaic language analysis the case of student academic papers. *VIEWS* 14.1. Retrieved August 24, 2005, from http://www.univie.ac.at/anglistik/ang_new/online_papers/views.html
- Jones, M. and Haywood, S. (2004). Facilitating the Acquisition of Formulaic Sequences: An Exploratory Study in an EAP Context. In N. Schmitt (ed.) *Formulaic Sequences*. Amsterdam: John Benjamins. 269-300.
- Kavaliauskiene, G. & Janulevieiene, V. (2004). Using the Lexical Approach for the Acquisition of ESP Vocabulary. *ITESLj*. Retrieved February 5, 2005, from <http://iteslj.org/Articles/kavaliauskiene-LA.html>.
- Khongphan, W. (2004). *Vocabulary Occurrences: An Analysis of Reading Passages in Comprehension 11*. Master's Project, M.A.(English). Bangkok: Graduate School Srinakharinwirot University. Photocopied.
- Kittikool, N. (2004). *An Analysis of Idiomatic and Literal Meanings of Phrasal verbs Used in Bangkok Post Front –Page Headlines*. Master Project, M.A.(English). Bangkok: Graduate School Srinakharinwirot University. Photocopied.
- Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal* 59 (4):322-332.
- Koya, T. (2004). Collocation Research Based on Corpora Collected from Secondary School Textbooks in Japan and in the UK. *Dialogue* 3: 7-18. Retrieved February 5, 2005, from http://homepage1.nifty.com/talk/dialogue/no03_2004/dialogue.03.kl.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.

- Lewis, M. (1996). Implications of a Lexical View of Language. In J. Willis & D. Willis, *Challenge and Change in Language Teaching*. Oxford: Heinemann. 10–16.
- (1997). *Implementing the Lexical Approach: putting theory into practice*. Hove: Language Teaching Publications.
- (1997). Lexical Implications of the Lexical Approach. In J.Coady & T.Huckin (eds.) *Second Language Vocabulary Acquisition*. Cambridge: CUP. 250-270.
- (2000). Language in the Lexical Approach. In M. Lewis, *Teaching Collocation: Further Developments in the Lexical Approach*. Hove: Language Teaching
- Lowe, C. (2003). *Lexical Approaches Now: The role of syntax and grammar*. Retrieved March 3, 2006, from <http://www.ihworld/ihjournal/CharlesLowe/asp>.
- Lyons, John. (2001). *Language and Linguistics*. Cambridge: CUP
- Lin, L.H.F. (2003). A Corpus Evidence: Existential Constructions by Chinese Learners of English. *The English Language Teacher*. 6 (3):279-290.
- McCarthy, M. (1998) Some Vocabulary Patterns in Conversation. In R. Carter & M. McCarthy (eds.). *Vocabulary and Language Teaching*. 181-200. London: Longman.
- (2002). What is an Advanced Level Vocabulary? In M.Tan (ed.) *Corpus Studies in Language Education*. Bangkok: IELE Press.15-29.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy, *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: CUP. 40–63.
- (1998) *Fixed expressions and idioms in English: a corpus-based approach*. Oxford: OUP.
- Nation I.S.P.(2000). *Learning Vocabulary in Another Language*. Cambridge: CUP.
- Nation I.S.P.(2003). Vocabulary. In D.Nunan (ed.) *Practical English Language Teaching*. Singapore: McGraw Hill. 43-56.
- Nation, I.S.P.& Newton, J. (1997). Teaching Vocabulary. In J.Coady & T.Huckin (eds.) *Second Language Vocabulary Acquisition*. Cambridge: CUP. 238-254.
- Nation, P. & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N.

- Schmitt & M. McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. 6-19.
- Nattinger, J.R. (1980). A Lexical Phrase Grammar for ESL. *TESOL Quarterly* 14 (3): 337-344.
- (1988). Some Current Trends in Vocabulary Teaching. In R. Carter & M. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman. 62-82.
- Nattinger, J.R. & DeCarrico, J.S. (1992) *Lexical phrases and language teaching*. Oxford: OUP.
- Oakey, D. (2000). A corpus-based study of the formal and functional variation in academic writing in social science, medical, and technical disciplines. In *BALEAP PIMs Reports*. Retrieved July 17, 2005 from <http://www.baleap.org.uk/pimreports/>
- O'Dell, Felicity. (1997). Incorporating vocabulary into the syllabus. In R. Carter & M. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman. 258-278.
- Pawley, A. & Syder, F.H. (1983) Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schmidt (eds). *Language and Communication*. London: Longman. 191–225.
- Peters, A.(1983). *The Units of Language Acquisition*. Cambridge: CUP
- Porto, M. (1998). Lexical Phrases and Language Teaching. *FORUM* 36/3. Retrieved February 5, 2005, from <http://exchanges.state.gov/forum/vols/vol36/no3/p22.htm>.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: CUP.
- Reed, A. (2006). *Simple Concordance Program 4.0*. Retrieved January 6, 2006, from <http://www.textworld.com/scp>
- Richards, J. & Schmidt, R.W. (1983). *Language and Communication*. London: Longman
- Schmitt, N. (1997). Vocabulary Learning Strategies. In N.Schmitt and M.McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: CUP. 199-227
- (2000). *Vocabulary in Language Teaching*. Cambridge: CUP.
- (2002). Using Corpora to Teach and Assess Vocabulary. In M.Tan (ed.) *Corpus Studies in Language Education*. Bangkok: IELE Press. 31-44

- Schmitt, N. & Carter, R. (2000). *Lexical Phrases in Language Learning*. Retrieved July 19, 2005, from <http://gean5.pfmb.unimb.si/eng/language/readings/lexicalphrases.htm>.
- Schmitt, N. & McCarthy, M. (1997) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: CUP.
- Simpson, R. & Mendis, D. (2003). A Corpus –Based Study of Idioms in Academic Speech. *TESOL Quarterly* 37 (3): 419-441.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. (2002). *Words and Phrases: Corpus Study of Lexical Semantics*. London: Blackwell Publishing.
- Stubbs, M. (1986). *Language Development, Lexical Competence and Nuclear Vocabulary*. Retrieved January 1, 2005, from <http://www.uni-trier.de/uni/fb2/anglistik/Projekte/tm>
- (2000). *Using Very Large Text Collections to Study Semantic Schemas: A Research Note*. Retrieved January 1, 2005, from <http://www.unitrier.de/uni/fb2/anglistik/Projekte/htm>.
- (2004). *On Very Frequent Phases in English: Distributions, Functions and Structures*. Retrieved January 1, 2005, from <http://www.unitrier.de/uni/fb2/anglistik/Projekte/stubbs/icame-2004>
- Tan, M. (2002). *Corpus Studies in Language Education*. Bangkok: IELE Press.
- (2002). Fixed Expressions, Prepositional Clusters and Language Teaching. In M. Tan (ed.) *Corpus Studies in Language Education*. 95-117. Bangkok: IELE Press.
- Thornbury, S. (1998). The Lexical Approach: A Journey Without Maps? *Modern English Teaching* 7 (4):7-13.
- (2002). *How to Teach Vocabulary*. Harlow: Longman.
- Thornbury, S. & Slade, D. (2006). *Conversation: From Description to Pedagogy*. Cambridge: CUP.
- Waring, R. (2002). Why should we build up a start-up vocabulary quickly. Retrieved April 20, 2005, from <http://www1.harenet.ne.jp/~waring/vocab/principles/early.htm>.

- Weinert, R. (1995). The Role of Formulaic Language in Second Language Acquisition: a review. *Applied Linguistics* 16 (2): 180-205.
- Willis, D. (1990). *The Lexical Syllabus: A New Approach to Language Learning*. London: Collins ELT.
- Willis, D. (1997). Lexical Phrases in Syllabus and Material Design. *The Language Teacher Online Journal*. Retrieved March 10, 2005, from <http://www.jalt-publications.org/tlt/files/97/sep/willis.html>.
- Willis, J. (1998). Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In B. Tomlinson, *Materials development in language teaching*. Cambridge: CUP. 44–66.
- Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics* 21 (4): 463–89.
- Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching* 32 (4): 213–31. Retrieved August 24, 2005, from <http://www.cardiff.ac.uk/ecap/clcr/flarn/wrayperkins.pdf>.
- (2002). Formulaic language in computer-supported communication: Theory Meets reality. *Language Awareness* 11(2):114-131. Retrieved August 24, 2005, from <http://www.channelview.publications.net>.
- Wray, A. & Cox, S., et.al. (2004). A formulaic approach to translation at the post office: reading the signs. *Language and Communication* 24: 59-75. Retrieved August 24, 2005, from <http://www.elsevier.com/locate/langcom.html>.
- Wray, A. & Perkins M.R. (2000). The functions of formulaic language: an integrated model. *Language and Communication* 20 (1):1–28. Retrieved August 24, 2005, from <http://www.elsevier.com/locate/langcom>.
- Yorio, C.A. (1980). Conventionalized language forms and the development of communicative competence. *TESOL Quarterly* 14 (4): 433–42.

APPENDICES

APPENDICES

Appendix A: Transcript Samples

Sample 1 Life's Experience

(Tapescript from English Upgrade 2, page 126.)

Ann: Hi, Maria. How're you doing?
 Maria: Oh, I don't know. Not so good, I guess.
 Ann: Really? What's the matter?
 Maria: It's my boyfriend. I forgot his birthday and
 now he won't talk to me.
 Ann: That's too bad. Have you called him to apologize?
 Maria: Well, I've tried to call, but he doesn't answer the phone.
 Ann: Have you e-mailed him?
 Maria: No, I haven't yet, but maybe I should.
 Ann: And maybe you should send him a romantic
 "I'm sorry" card.
 Maria: Mmm, yeah, that might help.
 Ann: And why don't you give him a bottle of nice
 aftershave or cologne?
 Maria: Hey, that's not a bad idea.
 Ann: Yep, that usually works for me.
 Oh, and one more good idea....
 Maria: Uh huh, what's that?
 Ann: You'd better write down his birthday on your calendar.
 Maria: Definitely! Well, thanks for the suggestions, Ann.
 Ann: No problem. Let me know what happens.

Sample 2: Making Complaints

(Tapescript from Person to Person 2, page 121)

A: Excuse me!
 B: Yes, Ma'am. What can I do for you?
 A: I have a problem with this coffeemaker,
 and I want my money back.
 B: Well, what seems to be the problem?
 A: The problem is that it doesn't work! I've
 only used it three times.
 B: I'd be happy to exchange it for you.
 A: Thank you but I'd like a refund, please.
 I invited six people to my place for dinner
 last night. After dinner, I plugged in the coffemaker,
 I put in the coffee and water, I turned it on and
 Nothing happened! I was very embarrassed. No I'd
 like my money back.
 B: Of course. A refund is no problem at all. I'm sorry
 you've had much trouble.
 A: That's OK. And thank you for your help.

Appendix B: Sample Analysis

Going on Holidays (Transcript from Interchange 2, page 121)

Julie: *That's great! The more the merrier!*

Nancy: *By the way*, where are we going to stay?

Julie: Oh. We can stay at my aunt and uncle's **beach house**. They have plenty of room and I'm sure they'll be happy to have guests. I'll call them tonight to let them know what time we're going to arrive. I guess, we'll leave **pretty early** .

There's a direct bus every morning at 5:00 A.M.

Nancy: *That's fine with me*. I think I'll be too excited to sleep!

Julie: I know! *And the best thing is*, we'll probably get there in time to spend a few hours *at the beach* after we get settled.

Legend: Multiword expressions are *italicized*, Polywords in **bold**

Appendix C: Samples of polyword found in the target English language university textbooks

1. **Phrasal verbs**- back up, broke into, bordered by, clean up, come in, dry out, faced with, filled up, find out, filled with, graduated from, got off, get into, grown in, hang up, lighten up, look for, look up, lying around, make up, made in, made from, moved into, pay back, pick up, put up, put on, swollen up, take some time off, turned into, turn off, wake up, work out
2. **Compounds** - alarm clock, answering machine, beach house, car accident, chocolate bar, credit card, computer course, educational system, horse back riding, household expenses, major points, mountain climbing, parking lot, parking zone, pain reliever, personal belongings, personal problem, phone number, e-mail address, reading light, registration form, room service, rural areas, vending machine, video game, soft drinks, summer camp, sunblock cream, speed limit, zip code
3. **Binomials** –brother and sister, dirtier and noisier, days and times, night and day, home and study, hot and humid, hundred and fifty, go and get, Ladies and gentlemen, mom and dad, aunts and uncles, napkins and towels, noon and three, now and then, pounding and pounding, sweet and sour, salt and pepper, work and school, up and down
4. **Collocations** - comfortable clothes, commit crimes, fantastic view, fresh air, get well paid, get a sunburn, high quality, hit the browser, keep fit, make a budget, make some money pleasant stay, soft music, strong glasses, take aspirin, take a walk, take a train, terrible headache, waste money, use the internet

Appendix D: Samples of Multi-word expression found in the target English language university textbooks

1. Fixed expressions

A bit boring
 And most important
 Any good suggestions?
 Don't mention it.
 How about you?
 How are you doing?
 I'd love to.
 I have to go.
 I guess so.
 I'm looking forward to it.
 I was wondering ..
 Just so-so
 Let me know what happened.
 Let me see.
 Oh, too bad.
 Tell me about it.
 That's not bad idea.
 That's great.
 Thanks a lot.
 See you later.
 Stuff like that.
 Wait a minute.
 What's your problem?/ matter?
 You can say that again.
 You can't miss it.

2. Semi-fixed expressions

Could you tell me _____?
 Do you have _____?
 Do you know where _____?
 Do you like _____?
 Do you live in _____?
 I guess that's kind of _____?
 I'm looking for _____.
Nice to meet you.
 You mean _____.
 So why don't we _____?
 Would you like _____?

3. Prepositional phrases

across from the exit, about two hours,
around eight, at the end, at the moment,
for a summer, for the summer, for two weeks,
in the country, in the program,
in front of, in the morning, in the meantime,
in two weeks, in a bank,
on the corner, on the weekends, onto the beach,
over there, next to me, up the hill

4. Idioms

the more the merrier.
work hard, play hard

VITAE

VITAE

Name: Jobeth A. Jang

Date of Birth: January 6, 1974

Place of Birth: Tawang, La Trinidad, Benguet, Philippines

Present Address: 50/631 Mhuban Sathaporn Soi 19 Baungyeetho, Thanyaburi

Pathumthani, 12130, Thailand

EDUCATIONAL BACKGROUND:

2008 Master of Arts degree in Teaching English as a Foreign Language

Srinakharinwirot University

Sukhumwit 23, Bangkok City, Thailand

2001 Certificate in Teaching English to Speakers of Other Languages

Projects For Asia, Baguio City, Philippines

1999 School of Frontier Missions – Trainer and Staff

Youth With A Mission, Los Banos, Laguna, Philippines

1995 Bachelor of Science in Physical Therapy

Virgen Milagrosa University

San Carlos City, Pangasinan, Philippines