

**A STUDY OF ACADEMIC WORDS IN SCIENCE NEWS:
A CORPUS-BASED STUDY**

**A MASTER'S PROJECT
BY
LUCKANA PANBANLAME**

**Presented in Partial Fulfillment of the Requirements
for the Master of Arts Degree in English
at Srinakharinwirot University**

October 2007

Copyright 2007 by Srinakharinwirot University

**A STUDY OF ACADEMIC WORDS IN SCIENCE NEWS:
A CORPUS-BASED STUDY**

AN ABSTRACT

BY

LUCKANA PANBANLAME

**Presented in Partial Fulfillment of the Requirements
for the Master of Arts Degree in English
at Srinakharinwirot University
October 2007**

Luckana Panbanlame. (2007). *A Study of Academic Words in Science News:*

A Corpus –Based Study. A Master's Project, M.A. (English). Bangkok:

Graduate School, Srinakharinwirot University. Advisory Committee:

Assist.Prof. Dr.Nitaya Suksaeresup, Mrs. Tuanta Laosooksri,

Dr. Walaiporn Chaya

The purpose of this study was to investigate the frequency of words in the AWL occurring in science news, to explore and to present 50 most frequent words in science news. The study also aimed to compare academic words of AWL between the science news in <http://www.bbc.co.uk> and <http://www.cnn.com>. The analysis tool in this study was the concordance program, and Coxhead (2000)'s Academic Word List (the AWL) used as the framework of this study.

The corpus of BBC science news contained 89,920 word tokens which the corpus of CNN science news contained 110,119 word tokens. The results of the study showed that the coverage of the AWL in BBC science news corpus and CNN science news corpus was 7% and 6% respectively. There were some differences between the corpus of BBC science news and CNN science news. Four hundred and twenty-two academic word families occur in both corpora but 108 academic word families occur either in the corpus of BBC science news or CNN science news.

การศึกษาคำศัพท์วิชาการในข่าววิทยาศาสตร์

บทคัดย่อ

โดย

ลักขณา พันธุ์บ้านแหลม

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต สาขาวิชาภาษาอังกฤษ
ตุลาคม 2550

ลักษณะ พันธุ์บ้านแหลม (2550). การศึกษาคำศัพท์วิชาการในข่าววิทยาศาสตร์ สารนิพนธ์ ศศ.ม.

(ภาษาอังกฤษ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

คณะกรรมการควบคุม: ผู้ช่วยศาสตราจารย์ ดร.นิตยา สุขเสรีทรัพย์

อาจารย์เดือนตา เลาสุขศรี อาจารย์ ดร.วไลพร ฉายา

การศึกษาคำศัพท์ครั้งนี้มีวัตถุประสงค์เพื่อสำรวจจำนวนคำศัพท์ทางวิชาการในข่าววิทยาศาสตร์

จากข่าวที่ปรากฏในเว็บไซต์ <http://www.bbc.co.uk> และ <http://www.cnn.com>. ว่ามีคำศัพท์

วิชาการใดที่พบมากเป็น 50 อันดับแรก และคำศัพท์ทางวิชาการที่พบในเว็บไซต์

<http://www.bbc.co.uk> และ <http://www.cnn.com>. มีความแตกต่างกันหรือไม่ การวิเคราะห์

จำนวนคำศัพท์ใช้โปรแกรมคอมพิวเตอร์ Concordance คำศัพท์วิชาการที่นำมาเป็นกรอบ

การศึกษาคำศัพท์ครั้งนี้คือกลุ่มคำศัพท์วิชาการ (The Academic Word List) จำนวน 570 คำ ซึ่งพัฒนา

โดยค็อกซ์ เฮด (2000) ผู้วิจัยได้รวบรวมข้อมูลจากข่าวอินเทอร์เน็ตเพื่อสร้าง Corpus จากเว็บไซต์

<http://www.bbc.co.uk> จำนวน 89,920 คำ และเว็บไซต์ <http://www.cnn.com> จำนวน 110,119

คำ ผลการศึกษาพบว่าคำศัพท์วิชาการในเว็บไซต์ <http://www.bbc.co.uk>. คิดเป็น 7 เปอร์เซ็นต์

และในเว็บไซต์ <http://www.cnn.com> คิดเป็น 6 เปอร์เซ็นต์ ผลจากการศึกษาคำศัพท์ทางวิชาการ

ทั้งสองเว็บไซต์ พบว่ามีความแตกต่างกัน คือ มีคำศัพท์ทางวิชาการจำนวน 422 คำที่พบในทั้งสอง

เว็บไซต์ และคำศัพท์ทางวิชาการจำนวน 108 คำที่พบในเว็บไซต์ใดเว็บไซต์หนึ่ง

The Master's Project Committee and Oral Defense Committee have approved this Master's Project as partial fulfillment of the requirements for the Master of Arts degree in English of Srinakharinwirot University.

Master's Project Advisor

.....*Nitaya Suksaeresup*.....

(Assistant Professor Dr. Nitaya Suksaeresup)

Chair of the Master of Arts Program in English

.....*Prapaipan Aimchoo*.....

(Dr. Prapaipan Aimchoo)

Oral Defense Committee

.....*Nitaya Suksaeresup*.....Chair

(Assistant Professor Dr. Nitaya Suksaeresup)

.....*Walaiporn Chaya*.....Reader

(Dr. Walaiporn Chaya)

.....*Tuanta Laosooksri*.....Reader

(Mrs. Tuanta Laosooksri)

This Master's Project has been approved as partial fulfillment of the requirements for the Master of Arts degree in English of Srinakharinwirot University.

.....*C. Pibulchol*..... Dean of the Faculty of Humanities

(Associate Professor Chaleosri Pibulchol)

October.....¹⁹....., 2007

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who gave me the possibility to complete my master's project. First of all, I would like to express my sincere gratitude and deep appreciation to Asst.Prof. Dr. Nitaya Suksaeresup, my master's project advisor, Ajarn Tuanta Laosooksri, and Ajarn Dr.Walaiporn Chaya, my committee members, for their invaluable guidance and useful encouragement and insightful advice which made this study possible.

I am also grateful to Asst.Prof. Dr. Tipa Thep-Ackrapong for her guidance, beneficial comments, willingness to help and kindness.

I am also grateful to all the lecturers in Western Languages Department, Srinakharinwirot University, for their patience to give me knowledge and experience.

I wish to thank to my classmates in this program for their help in difficult times and untiring encouragement.

Lastly, I would like to express my deepest gratitude to my parents and my grandmother, who have always given me assistance, understanding and encouragement.

Luckana Panbanlame

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION.....	1
Background.....	1
Research Questions.....	3
Scope of the Study.....	3
Significance of the Study.....	3
Definition of Terms.....	4
Overview of the Study.....	5
2 REVIEW OF LITERATURE.....	6
Vocabulary Learning.....	6
Word Groups.....	7
The Academic Word List (The AWL).....	8
Corpus-based Study.....	11
Related Literature of Academic Words.....	12
3 METHODOLOGY.....	15
Analytical Tool.....	15
Analytical Framework.....	15
The Concordance Program (CP).....	16
Procedure.....	16
Data Collection.....	16
Corpus Creation.....	16
Data Analysis.....	16

TABLE OF CONTENTS (CONTINUED)

Chapter	Page
4 FINDINGS.....	19
5 CONCLUSION AND DISCUSSION.....	36
Conclusion.....	36
Discussion.....	38
Applications of the Study.....	39
Limitations of the Study.....	40
Suggestions for Further Studies.....	40
REFERENCES.....	45
APPENDIX.....	49
VITAE.....	61

LIST OF TABLES

Table	Page
1 Subject areas of the Academic Corpus	10
2 Academic word families, word types, word tokens and the coverage of the AWL in the corpus of BBC science news.....	19
3 Word families, word types, word tokens and the coverage of the AWL in the corpus of CNN science news.....	20
4 The 50 most frequent word families from the corpus of BBC science news.....	21
5 The 50 most frequency word families from the corpus of CNN science news.....	24
6 A comparison of the BBC science news and CNN science news corpora	27
7 Word families in the AWL appeared both in the corpus of BBC science news and CNN science news.....	28
8 Word families appeared either in the corpus of BBC science news or CNN science news	33

LIST OF FIGURES

Figure	Page
1 A sample of the Concordance Program.....	15
2 A sample of academic words in http://www.bbc.co.uk	35
3 A sample of academic words in http://www.cnn.com	37

Chapter 1

Introduction

1.1 Background

Learning vocabulary is one of the important factors in learners' language development. Donley and Reppen (2001) agreed that students are not able to develop their language because of their vocabulary shortage. In the class room, the students understand class discussion, lectures and texts because of their vocabulary knowledge. Furthermore, vocabulary knowledge is related to reading comprehension. The objective of reading is to build text meaning. Learners read English effectively because they have enough vocabulary knowledge.

There are a large number of vocabularies in English. In 1990, Goulden, Nation, and Read counted the number of word families in Webster's Third New International Dictionary which is the largest dictionary of English (Schmitt 2000). They found that there were 54,000 word families. Therefore, it is impossible that second language learners can remember all English words. Some researchers classified the words into groups. For this reason, teachers and learners can focus on what vocabulary they need to learn.

Actually, studying academic words is very valuable for research work because students always read academic text. Academic words are high level vocabulary. Therefore, teachers have to carefully plan how to teach academic words effectively and apply suitable activities to support learner's vocabulary learning. Generally, academic words do not occur in daily life, so learners do not have an opportunity to receive or produce academic words in listening, speaking, reading and writing. As a result, they do not develop the fluency of using words (Nation 2001). Schmitt (2000) points out that if learners are encouraged to have experience of words, learners increase the number of their

words as well. However Nation, P (1994) argues that it is unnecessary to have a great number of vocabularies. In addition, it is more essential that learners be fluent in terms of use. However, there are considerable English words; therefore, it is impossible to know all English words. Most importantly, teachers should focus on specialized words which are utmostly favorable to students. Learning at higher education level, students need to learn academic words so that they can understand academic texts more easily. Finally, both English as the first language learners and the second language learners face problems of academic words. Santos (2003) investigated academic vocabulary knowledge of U.S. language minority community college students. The result showed that Asian students in ESL classes and language minority students had the same problem of academic words because they had no chance using words in an academic field.

This study aimed to investigate the frequency of words in the AWL which appear in the science news, to explore 50 most frequent words in science news and to compare the academic words in AWL between the science news in <http://www.bbc.co.uk> and <http://www.cnn.com>. The researcher chose to analyze the academic words in <http://www.bbc.co.uk> and <http://www.cnn.com> because these two websites present the most famous news of the world. [Http://www.bbc.co.uk](http://www.bbc.co.uk) is from the United Kingdom whereas <http://www.cnn.com> is from the United States of America. Moreover, learning language on the internet has considerable advantages. The second language learners encounter natural language and authentic language. The teachers can use the information on the internet to teach students. It is meaningful language task. The news on the Internet is current and frequently updated. Learners can surf the Internet every time and pay attention what topics they are interested in. It is an independent learning. Learners or audiences get advantages from the Internet.

1.2 Research Questions

This study explored the following questions:

1. What is the frequency of the academic words in the AWL in <http://www.bbc.co.uk> and <http://www.cnn.com>?
2. What are the 50 most frequent academic words in science news?
3. Are there any differences of the words in the AWL of science news between <http://www.bbc.co.uk> and <http://www.cnn.com>, and if they are what are they?

1.3 Scope of the study

Science news was collected from <http://www.bbc.co.uk> and <http://www.cnn.com> from 1 April 2006 to 1 April 2007. The analysis was in the following aspects:

1. The number of Coxhead's AWL words in the corpus of science news.
2. The top 50 AWL words in the corpus of science news.
3. A comparison of academic words in science news between <http://www.bbc.co.uk> and <http://www.cnn.com>.

1.4 Significance of the study

The aim of this study was to explore the number of academic words in science news from <http://www.bbc.co.uk> and <http://www.cnn.com>. The implications of this study are as follows:

1. An academic word list is useful for setting vocabulary goals for language courses.
2. It is an authentic teaching aid of ESP and EAP texts that can be used to develop the students' reading comprehension of the science text.

1.5 Definitions of Terms

1. Academic words

Academic words are found in various kinds of academic texts from 3.5 million running words. The Academic Word List (The AWL) was created by Coxhead in 2000. The AWL consists of 570 words which occur frequently in a very wide range of academic texts.

2. Tokens

It is a counting unit of word occurrence in a spoken or written text. For example, the sentence, *when I was young, I wanted to be a teacher*. This sentence consists of 10 tokens.

3. Word types

It is a counting unit of different words occurring in the text. From the above sentence, it consists of 10 tokens but it has 9 word types.

4. Word families

A word family consists of a head word, its inflections, and its derivation. There are plural, third person singular present tense, past tense, past participle, comparative, superlative and affixes (-able, -ible, -r, -er, -ic, -ish, -less, -ly, -ness, -y, -al, -ation, -ess, -ful, -ism, -ist, -ty, -ize, -ment, in-, non-, and un-). For example, *conception, concepts, conceptual, conceptualization, conceptualize, conceptualized, conceptualizes, conceptualizing* and *conceptually*.

5. Corpus

A corpus is a collection of both speech and written texts. A corpus is analyzed to identify either characteristics or variety of a language (Sinclair 1991).

6. Concordance

Concordance refers to a list of all words that are used to analyze a specific text. Concordance includes language patterns in the text. It studies word frequency, word lists, grammar features, discourse and genre. (Sinclair 1991).

1.6 Overview of the study

This study aimed to investigate the frequency of words in the AWL which appeared in science news, explore the 50 most frequent words in science news, and compare the academic words in the AWL between the science news in <http://www.bbc.co.uk> and <http://www.cnn.com>. The review of literature in chapter 2 provided information about vocabulary acquisition, word groups, the Academic Word List (The AWL), corpus-based study, related literature of academic words. Chapter 3 presented the methodology of research including the analytical tool and procedures. Chapter 4 presented the results of the study. Chapter 5 presented the conclusion and discussion and applications of the study, limitations of the study and suggestions for further studies.

Chapter 2

Review of Literature

This chapter presents literature and relevant research. It consists of vocabulary acquisition, word groups, academic words, a corpus-based study and related research.

2.1 Vocabulary Acquisition

2.1.1 Vocabulary Learning

The study of vocabulary learning is fundamental of relevant studies of words. Nation, P (1994) and Schmitt (2000) agree that it is essential to study how vocabulary is learned and used by learners. Vocabulary learning consists of explicit and incidental vocabulary learning. Both explicit and incidental vocabulary learning are useful for second language learners.

Explicit vocabulary learning involves the teaching of vocabulary. It emphasizes study of words. Learners are encouraged to remember target words by matching the word concept and imagination. Consequently, learners concentrate on words which are found in the text. They are bored with vocabulary teaching activities. Schmitt (2000) concludes the concept of explicit vocabulary learning as follows. The teachers present a large concept of vocabulary then combine new words with the old ones. Besides, learners can see word usage from sample sentences. They are encouraged to learn vocabulary more deeply through the use of objects and word meaning. There are many kinds of vocabulary teaching techniques. However, there is a weakness of explicit learning vocabulary teaching. Learners are more confused when they learn more than two relevant words at the same time. Similarly, it is very difficult for learners to understand the concept of a word because many English words have many substantial meanings.

On the other hand, learners increase vocabulary using incidental vocabulary learning. This learning focuses on language use. Incidental vocabulary learning is used for communicative purposes. It considers applying words in daily life following function words. Learners are familiar with words. In addition to reading, the more learners read considerable text, the more they know words. Learners learn words gradually because they always see words in the text. However, it takes a long time to remember words. Obviously, the learning of this process is quite slow. Furthermore, learners' long term memory is from understanding the whole comprehension of word meaning. Nation, P (2001) and Coady (2002) claimed that learners understand the text even though they do not know the word meaning. It is interesting how learners extend vocabulary from reading a text. That is, reading comprehension deals with the reader's capacity and the text. The reader tries to understand the unknown words with deep meaning. Sometimes there is no cue to explain words. It is a difficult task for the reader to guess the meaning of words. Without a cue, ESL readers probably guess wrong meaning rather than guess correct meaning (Coady 2002). Incidental vocabulary learning is beneficial for reading as follows. Firstly, learners pay attention to new words which are clearly explained in the context. Secondly, learners read a lot of texts so that learners increase a large number of vocabularies. Finally, learners understand the context easier because of word categories.

2.1.2 Word Groups

(Nation, P 2001) classifies words in reading texts into four main groups: high frequency words, academic words, technical words and low frequency words.

High frequency words

High frequency words are words that are widely used. High frequency words can be found frequently in ordinary texts. Furthermore, they also include the language use in different situations. West's (1953) grouped high frequency words which are generally known as General Service List (GSL) which comprises 2,000 word families. These word families are divided into two groups; function words and content words. The sample of function words are *a, some, two, because* and *to*. Similarly, content words are nouns, verbs, adjectives, and adverbs such as *agree, time, slow* and *sometimes*. It is stated that 2,000 word families are found 87% of the running words in formal written texts and more than 95% of the words in informal spoken texts (Nation 1994).

Low frequency words

The occurrence of low frequency words is hardly found in any texts. There are approximately 100,000 word families in *Webster's Third International Dictionary*. According to language teaching, it is unnecessary to learn low frequency words. However, if learners would like to increase the vocabulary size, learners can learn low frequency words outside the class by themselves.

Technical words

Technical words are specialized words used in a particular field. It is difficult to tell a certain number of technical words and explain definitions of technical words. Some words are also high frequency words but their meanings are technical terms. For example, *accused* (noun), *offer* and *reconstruction* (of a crime) are both law technical terms and high frequency words.

Academic words

Academic words are found in various kinds of academic texts from 3.5 million running words such as *area*, *data*, *issue* and *require*. Academic words are created by Coxhead in 2000. Academic Word List consists of 570 words in academic texts; however, the AWL is not the most frequent words of English.

2.1.3 The Academic Word List (The AWL)

The AWL was published in 2000 by Averil Coxhead of Victoria University of Wellington, New Zealand. Coxhead's Academic Word List was chosen to investigate academic words in academic texts. The AWL is not the first word list to analyze vocabulary. In the past, several researchers investigated vocabulary in academic texts. Initially, *General Service List (GSL)* was created by West's in 1953. The purpose of GSL is to analyze English scientific vocabulary. After that, *the University Word List (UWL)* of Xue and Nation (1984, also in Nation, 1990) consists of 800 word families. It aims to enhance university students' vocabulary development. UWL had 8.5 % coverage of academic texts. Then UWL was replaced by the Academic Word List (The AWL) of Coxhead in 1998. There are two reasons why UWL was replaced. Firstly, the AWL is shorter than UWL. Specifically, the list of 570 word families is based on 3,500,000 token corpus of academic English. Secondly, the AWL has better coverage than the UWL.

The purpose of Coxhead's study was to develop and evaluate a new academic word list. There were two research objectives to develop the AWL; namely, to survey academic words occurrence found in different academic sources and to differ academic words frequencies in arts, commerce, law and science texts. There were four objectives to evaluate the AWL; namely, to study the AWL's coverage of the words in the Academic corpus, to explore the frequency of word in another academic corpus, and the frequency of

words in the AWL occurring in nonacademic texts and to compare the AWL and UWL (Xue & Nation 1984).

With regard to the academic corpus, 414 academic texts contained 3,513,330 tokens (running words) which are equivalent to 70,377 word types. Academic word corpus was presented in different kinds of academic texts as follows; 158 articles from academic journal, 51 edited academic journal articles from the websites, 43 university textbooks, 115 texts from the Learned and Scientific sections of the Wellington, Corpus of Written English, Brown corpus and Lancaster – Osto Berger corpus (LOB), 13 books from Micro Concord academic corpus and 2 university psychology laboratory manuals. Furthermore, the academic corpus was divided into four main disciplines: arts, science, law and commerce. Each discipline consists of seven subject areas. Therefore, there are twenty-eight subject areas.

Table 1 Subject areas of the Academic Corpus

Arts	Commerce	Law	Science
Education	Accounting	Constitutional Law	Biology
History	Economics	Criminal Law	Chemistry
Linguistics	Finance	Family Law and Medico-Legal	Computer Science
Philosophy	Industrial Relations	International Law	Geography
Politics	Management	Pure Commercial Law	Geology
Psychology	Marketing	Quasi-Commercial Law	Mathematics
Sociology	Public Policy	Rights and Remedies	Physics

Table 1 presents the twenty – eight subject areas of four disciplines, namely arts, science, law and commerce.

Specialized occurrence, range and frequency were important criteria to analyze the Academic corpus. Specialized occurrence referred to the word families not included in the 2000 words in GSL (West's 1953). Significantly, a member of a word family occurred at least 10 times in each four main disciplines and 15 or more in 28 subject areas. It was

called range. Range points out learners' capacity of using academic words. If the percentage was too low, the content and the form of the academic text were quite low. Frequency identified word occurrence which was found at least 100 times in the Academic corpus.

Coxhead's findings showed the occurrence of academic words. Some most frequent word families in the AWL were *analyze, concept, data* and *research*. The least frequent words were *convince, notwithstanding, going, persist* and *whereby*. Furthermore, there were differences in occurrence of words across four disciplines. That is academic words were found in commerce, law, arts and science calculated 12%, 9.4% and 9.1 % respectively. According to evaluation of the AWL, the AWL accounted for 10 % of the running words in the Academic corpus. In contrast, the AWL accounted for 1.4 % of the running words in fictions. It indicated that most academic words were found in academic texts rather than non academic texts. Finally, Coxhead compared the AWL with the UWL. In addition to size, the UWL contained 836 word families while the AWL contained exclusively 570 word families. The AWL's coverage of academic corpus was 10 % whereas the UWL's coverage was 9.8 %. In short, the AWL is more appropriate than the UWL in terms of the size and the coverage.

2.2 Corpus-based study

A corpus-based study is not new knowledge of linguistics. It has been available nearly 30 years. Several researchers investigated the use of corpus in the classroom.

In addition to the characteristics of a corpus-based study, Biber, D and Conrad, S (2001) define the characteristics of a corpus based analyses. Namely, it is a study of a collection of texts both in spoken and written forms. The text is analyzed by a computer program which is called concordancer. The corpus is classified into two types. Firstly,

the target language corpus is employed in the classroom. Secondly, a target language and native language corpus dealt with comparing the target language to the native language. It was used in translation.

A corpus is analyzed by a computer program called a concordancer. Concordancer is a software tool which facilitates the researchers. It is also easy to use. Additionally, it computes all occurrences all the words in the corpus and creates a frequency list. A frequency list is arranged in the alphabetical order or frequency order. Furthermore, the researchers designate words they would like to study. Concordancer illustrates specific words in the corpus.

2.3 Related Literature of Academic Words

Nakprakhon (2005) surveyed academic words in the political reports at www.nationmultimedia.com. The findings indicated that academic word families were found 90 % of the AWL. Words in the AWL covered 5.7 % of the total words in the corpus. The sample academic words in the political reports are *constitution*, *prime*, *policy* and *economy*.

Khongphan (2004) studied words frequency in an English reading textbook of Mattayom 5. It was found that the reading text book contained academic words 3.4 % of total running words in the corpus.

Santos (2003) investigated academic words knowledge of language minority community college students. The purpose of the study was to point out how many students knew academic words and how well students knew academic words. The result showed that students who knew numerous academic words were good at using academic words as well.

Poonpon (2002) analyzed academic words in the AWL of the intensive and extensive course materials provided for first year and second year science students at Mahidol and Khon Kaen University. It was found that both Mahidol and Khonkaen University second year students' course materials contained more additional academic words in the AWL than first year students' course materials. In Mahidol University's course material, there were 328 academic word families in the first year students' text and 389 academic word families in the second year students' text. With regard to Khon Kaen University, 201 academic word families occurred in the first year students' course material and second year students' material contained 393 word families. It was clearly shown that 70 academic word families occurred more than ten times. Some of these words are *energy*, *paragraph*, *process* and *compute*.

Kaewphanngam (2002) compared the first and the third year Silpakorn University's comprehension of technology and academic vocabulary. The study revealed that students comprehended academic words more than technical words. The top five most frequent academic words which were found in psychology text were *research*, *participants*, *psychology*, *analysis* and *process* respectively. The top five least frequent academic words were *themes*, *suppose*, *route*, *removed* and *plus*.

In short, this chapter presented vocabulary acquisition, word groups, the Academic Word List (The AWL), corpus-based study, and related literature of academic words. Vocabulary acquisition consisted of vocabulary learning, word groups, the Academic Word List (The AWL). Vocabulary learning consisted of explicit and incidental learning. Word groups were categorized by Nation, P (2001) into 4 groups: high frequency words.

low frequency words, technical words and academic words. Next, the Academic Word List (The AWL) created by Averil Coxhead was clearly described so as to give information about frequency of words in academic and nonacademic texts. Then the corpus-based study was defined and classified. This type of study involves analyzing a collection of texts by a computer program called a concordance in order to compute occurrence of words in the corpus and creates a frequency list in alphabetical order and frequency order. Lastly, research studies academic words in different disciplines conducted in Thailand and abroad were surveyed and reported.

Chapter 3

Methodology

This chapter presents the methodology of research divided into two parts: analytical tool and the procedures.

3.1 Analytical Tool

3.1.1 Analytical framework

Coxhead's Academic Word List (AWL) was employed to analyze the data. The AWL contained 570 words which are selected from 3.5 million running words. The AWL was divided into four areas. The four areas deal with arts, science, law and commerce. Each area is sub divided into seven general areas; therefore, the overall areas are 28 subject areas.

3.1.2 The Concordance Program

The concordance program is the software program to analyze the corpus in terms of word frequencies, word lists, and word concurrencies.

Figure 1 A sample of the Concordance Program

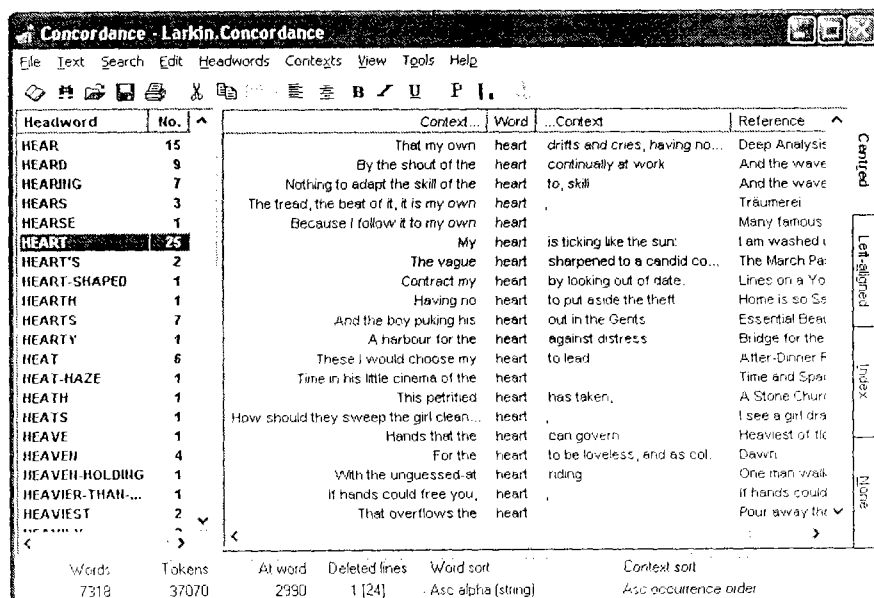


Figure 1 illustrates a sample of the concordance program. It demonstrates word lists, word frequencies, and different usage of a word.

3.2 Procedure

In order to meet the objectives, the following procedures were performed.

3.2.1 Data Collection

Data were collected from science news at <http://www.bbc.co.uk> and <http://www.cnn.com>. on every Monday, Wednesday and Friday from 1 April 2006 to 1 April 2007. Totally 320 news were selected from two websites.

3.2.2 Corpus Design

The steps in creating a corpus were as follows:

- a) Science news was searched from <http://www.bbc.co.uk> and <http://www.cnn.com>, and then was copied separately into Microsoft Word.
- b) Headlines and the body of the news were copied in to a file except dates and months.
- c) The texts were gathered into science news from <http://www.bbc.co.uk> and <http://www.cnn.com>. The corpus was fully completed.

3.2.3 Data Analysis

To answer the research questions, the concordance program and Coxhead's Academic Word List (AWL) were used to analyze the data. The procedure was carried out as follows:

Research Question 1: What is the frequency of the words in the AWL in

<http://www.bbc.co.uk> and <http://www.cnn.com>. ?

- a) The running words of the corpus were analyzed by employing the option Make Full Concordance
- b) The total running words of the corpus were shown in the program.
- c) The total AWL words were calculated using the Concordance's Pick List Tool. This tool included the total words of the AWL, headwords and their families.
- d) The percentage of AWL was calculated as follows:

$$\text{The percentage of the AWL} = \frac{\text{total AWL running words} \times 100}{\text{total running words in the text}}$$

Research Question 2: What are the 50 most frequent academic words in science news?

In order to answer this question, the frequency of each word was explained as follows:

- a) The frequency of each word in the AWL was compared to the findings of the first research question.
- b) The top 50 most academic words were examined and stored.
- c) The top 50 most academic words were reported in tables.

Research Question 3: Are there any differences of the words in the AWL of science news between <http://www.bbc.co.uk> and <http://www.cnn.com>, and if they are what are they?

The procedures were as follows:

- a) There were two corpora; namely corpus of science news at <http://www.bbc.co.uk> and <http://www.cnn.com>. Each corpus was analyzed separately.
- b) The running words of the two corpora were analyzed by employing the option Make Full Concordance.
- c) The total running words of the corpus were shown in the program.
- d) The total AWL words were calculated using the Concordance's Pick List Tool.
- e) The percentage of AWL was calculated as follows:

$$\text{The percentage of the AWL} = \frac{\text{total AWL running words} \times 100}{\text{total running words in the text}}$$

- f) The percentage of the AWL in <http://www.bbc.co.uk> compared with that in <http://www.cnn.com>.

In short, this chapter illustrated analytical tool and the procedures. The analytical tool consisted of the analytical framework and the concordance program. Coxhead's Academic Word List (the AWL) was used in the analytical framework. Two corpora of science news were analyzed by the concordance program to explore word lists and word frequencies. The procedures were performed to answer three research questions as follows: Firstly, what is the frequency of the academic words in the AWL in <http://www.bbc.co.uk> and <http://www.cnn.com>. ? Secondly, what are the 50 most frequent academic words in science news? Third, are there any differences of the words in the AWL of science news between <http://www.bbc.co.uk> and <http://www.cnn.com>, and if they are what are they?

Chapter 4

FIDINGS

This chapter presents the frequency of words in the AWL in science news. The words were analyzed into the AWL word families, the AWL word types, the AWL word tokens, total word types in the corpus, total word tokens in the corpus and the AWL coverage (%). The 50 most frequent used academic words in the AWL which occurred in science news were listed. Finally, the corpora from two websites were compared to examine the difference of the words in the AWL.

Table 2 Academic word families, word types, word tokens and the coverage of the AWL in the corpus of BBC science news

Type of words	BBC science news
The AWL word families	476
The AWL word types	1,127
The AWL word tokens	6,102
Total word types in the corpus	11,209
Total word tokens in the corpus	89,920
The AWL coverage (%)	7

Table 2 shows word families, word types, word tokens, total word types, total word tokens and the AWL coverage in the corpus of BBC science news. The results showed that the corpus of BBC science news contained 89,920 word tokens or running words. These words were classified into 1,127 word types (10 % of the total word types) and 476 word families (84 % of the AWL). The word families in the AWL covered approximately 7% of the text.

Table 3 Word families, word types, word tokens and the coverage of the AWL in the corpus of CNN science news

Type of words	CNN science news
The AWL word families	476
The AWL word types	1,174
The AWL word tokens	6,314
Total word types in the corpus	13,090
Total word tokens in the corpus	110,119
The AWL coverage (%)	6

Table 3 shows word families, word types, word tokens, total word types, total word tokens, and the AWL coverage in the corpus of CNN science news. It contained 110,119 word tokens or running words. These words were classified into 1,174 word types (9 % of the total word types) and 476 word families (84 % of the AWL). The word families in the AWL covered approximately 6% of the text.

Table 4 The 50 most frequent word families from the corpus of BBC science news

Rank	Word Families	Frequency	Percent (%) of text coverage
1	research	256	4.20
2	found	157	2.57
3	team	132	2.16
4	energy	128	2.10
5	environment	117	1.92
6	area	89	1.46
7	differentiate	89	1.46
8	data	87	1.43
9	process	69	1.13
10	project	68	1.11
11	technology	68	1.11
12	globe	64	1.05
13	impact	60	0.98
14	respond	60	0.98
15	region	59	0.97
16	device	56	0.92
17	evident	56	0.92
18	image	55	0.90
19	currency	53	0.87
20	journal	50	0.82
21	similar	48	0.79
22	site	46	0.75
23	target	45	0.74
24	structure	44	0.72
25	analyze	43	0.70

Table 4(continued)

Rank	Word Families	Frequency	Percent (%) of text coverage
26	create	43	0.70
27	major	42	0.69
28	consume	41	0.67
29	institute	39	0.64
30	design	38	0.62
31	expert	38	0.62
32	affect	37	0.61
33	complement	37	0.61
34	previous	37	0.61
35	vehicle	37	0.61
36	issue	36	0.59
37	predict	35	0.57
38	identify	34	0.56
39	indicate	34	0.56
40	physical	34	0.56
41	monitor	33	0.54
42	reveal	33	0.54
43	survey	33	0.54
44	conclude	29	0.48
45	display	29	0.48
46	benefit	28	0.46
47	considerable	28	0.46
48	equip	28	0.46
49	involve	28	0.46
50	potential	28	0.46
50	source	28	0.46

Table 4 shows the 50 most frequent academic words in the corpus. The 50 most frequent started from 28 up to 256 times. The highest frequency of occurrence was the word *research* (256 word occurrences or 4.20 % of text coverage). The 10 highest frequency of occurrence were the words *research*, *found*, *team*, *energy*, *environment*, *area*, *differentiate*, *data*, *process* and *project* respectively. The 10 lowest frequency of occurrence were the words *source*, *potential*, *involve*, *equip*, *considerable*, *benefit*, *display*, *conclude survey*, *reveal*, *monitor* and *physical* respectively. Some words had the same frequency such as *benefit*, *source*, *potential*, *involve*, *equip*, and *considerable*.

Table 5 The 50 most frequency word families from the corpus of CNN science news

Rank	Word Families	Frequency	Percent (%) of text coverage
1	research	216	3.42
2	percent	126	2.00
3	found	123	1.95
4	technology	106	1.68
5	environment	100	1.58
6	create	98	1.55
7	nuclear	96	1.52
8	energy	93	1.47
9	project	84	1.33
10	team	82	1.30
11	site	81	1.28
12	currency	76	1.20
13	area	62	0.98
14	design	55	0.87
15	estimate	55	0.87
16	involve	53	0.84
17	issue	53	0.84
18	secure	51	0.81
19	domain	47	0.74
20	major	47	0.74
21	respond	47	0.74
22	virtual	47	0.74
23	decade	45	0.71
24	differentiate	44	0.70
25	expert	44	0.70

Table 5(continued)

Rank	Word Families	Frequency	Percent (%) of text coverage
26	potential	43	0.68
27	complement	42	0.67
28	data	42	0.67
29	considerable	41	0.65
30	challenge	39	0.62
31	predict	38	0.60
32	invest	36	0.57
33	institute	35	0.55
34	similar	35	0.55
35	community	34	0.54
36	identify	34	0.54
37	monitor	34	0.54
38	react	34	0.54
39	target	34	0.54
40	final	33	0.52
41	physical	33	0.52
42	focus	32	0.51
43	fund	32	0.51
44	medical	32	0.51
45	construct	31	0.49
46	equip	31	0.49
47	generate	31	0.49
48	journal	31	0.49
49	normal	31	0.49
50	panel	31	0.49

Table 5 shows the 50 most frequent academic words in the corpus. The 50 most frequent words started from 31 up to 216 times. The highest frequency of occurrence was the word *research* (216 word occurrences or 3.42% of text coverage). The 10 highest frequency of occurrence were the words *research*, *percent*, *found*, *technology*, *environment*, *create*, *nuclear*, *energy*, *project*, and *team* respectively. The 10 lowest frequency of occurrence were the words *panel*, *normal*, *journal*, *generate*, *equip*, *construct*, *medical*, *fund*, *focus*, *physical* and *final* respectively. Some words had the same frequency such as *construct*, *equip*, *generate*, *journal*, *normal*, and *panel*.

Table 6 A comparison of the BBC science news and CNN science news corpora

Type of words	BBC science news	CNN science news
The AWL word families	476	476
The AWL word types	1,127	1,174
The AWL word tokens	6,102	6,314
Total word types in the corpus	11,209	13,090
Total word tokens in the corpus	89,920	110,119
The AWL coverage (%)	7	6

Table 6 shows types of words in the corpora of BBC science news and CNN science news. The total word tokens in the corpus of BBC science news were fewer than the total word tokens in the corpus of CNN science news. Specifically, the AWL word types and the AWL word tokens of BBC science news were fewer than the corpus of CNN science news. The two corpora of science news had the equal amount of the AWL word families. There were 476 the AWL word families. Finally, the AWL coverage of the corpus of BBC science news was one percent higher than that of CNN science news.

Table 7 Word families in the AWL appeared both in the corpus of BBC science news and CNN science news

abandon	assume	commission
academy	assure	commit
access	attach	commodity
accompany	attitude	community
accumulate	attribute	compatible
accurate	authority	compile
achieve	automate	complement
acknowledge	available	complex
acquire	aware	component
adapt	assemble	compound
adjust	assess	comprehensive
administrate	assign	comprise
adult	assist	compute
affect	bond	conceive
aggregate	brief	concentrate
aid	capable	concept
allocate	capacity	conclude
alter	category	conduct
analogy	challenge	confirm
analyze	channel	conflict
annual	chemical	consequent
anticipate	circumstance	considerable
apparent	cite	consist
appreciate	civil	constrain
approach	classic	construct
appropriate	coincide	consult
approximately	collapse	consume
area	colleague	contact
aspect	comment	contemporary

Table 7(continued)

converse	despite	enormous
convert	detect	ensure
convince	devote	environment
cooperate	differentiate	equip
context	dimension	equivalent
contract	displace	erode
contribute	display	error
controversy	dispose	establish
core	distinct	estate
corporate	distort	estimate
correspond	distribute	ethic
couple	diverse	evaluate
create	document	eventual
credit	domain	evident
criteria	domestic	evolve
crucial	dominate	exceed
culture	draft	exclude
currency	drama	exhibit
cycle	duration	expand
data	dynamic	expert
debate	economy	exploit
decade	element	export
decline	eliminate	expose
define	emerge	external
definite	emphasis	extract
demonstrate	enable	factor
deny	encounter	federal
depress	energy	file
derive	enforce	final
design	enhance	finance

Table 7(continued)

flexible	incidence	issue
focus	income	item
formulate	incorporate	job
forthcoming	indicate	journal
foundation	individual	label
found	induce	labour
framework	inevitable	layer
function	infer	lecture
fund	inherent	legal
fundamental	initial	legal
generate	initiate	legislate
globe	injure	liberal
goal	innovate	license
grade	input	link
grant	insight	locate
guarantee	inspect	logic
guideline	instance	maintain
hierarchy	institute	major
highlight	integrate	manipulate
identical	intelligence	margin
identify	intense	mature
ignorance	interact	maximize
illustrate	internal	mechanism
image	interpret	media
impact	interval	medical
implement	intervene	medium
implicate	invest	method
imply	investigate	migrate
impose	involve	military
incentive	isolate	minimal

Table 7(continued)

minimize	obtain	prior
mode	obvious	priority
modify	occupy	proceed
monitor	occur	process
mutual	odd	professional
network	offset	prohibit
neutral	option	project
nevertheless	orient	protocol
nonetheless	outcome	psychology
normal	output	publish
notion	overall	quote
nuclear	panel	radical
obtain	partner	random
obvious	perceive	range
occupy	percent	rational
occur	period	react
odd	phase	refine
offset	phenomenon	regime
option	philosophy	region
orient	plus	regulate
outcome	policy	release
output	pose	relevant
overall	positive	rely
panel	potential	remove
partner	precise	require
perceive	predict	research
percent	previous	revolve
period	primary	resource
phase	prime	respond
phenomenon	principal	restore

Table 7(continued)

restrict	straightforward	transmit
reveal	strategy	transport
reverse	stress	trend
revise	structure	trigger
revolution	style	ultimate
role	submit	underline
route	subsequent	unique
scenario	substitute	vary
schedule	successor	vehicle
scheme	sufficient	version
section	survey	via
sector	survive	virtual
secure	sustain	visible
seek	symbol	vision
select	target	volume
sequence	task	voluntary
series	team	welfare
sex	technical	whereas
shift	technique	widespread
significant	technology	
similar	text	
simulate	theme	
site	theory	
so-called	thereby	
sole	topic	
source	trace	
specific	tradition	
sphere	transfer	
stable	transform	
statistic	transit	
status		

Table 7 shows that there were 422 word families in both BBC science news and CNN science news.

Table 8 Word families appeared either in the corpus of BBC science news or CNN science news

accommodate	convene
advocate	deduce
albeit	deviate
alternative	device
amend	edit
amend	entity
annual	feature
approximate	fee
attain	fluctuate
behalf	format
benefit	furthermore
bias	generation
bulk	hence
chart	hypothesis
clause	ideology
code	immigrate
coherent	index
communicate	infrastructure
compensate	inhibit
confer	insert
conform	inspect
consent	instruct
constant	integrate
constitute	integrity
contradict	likewise
contrast	Mental

Table 8(continued)

minimum	reject
ministry	relax
minor	reluctant
motive	reside
norm	restrain
notwithstanding	retain
objective	revenue
overlap	scope
overseas	somewhat
paradigm	specify
parallel	subside
parameter	sum
participate	summary
passive	suspend
persist	tape
perspective	temporary
portion	tense
practitioner	thesis
preliminary	undergo
presume	undertake
principle	uniform
promote	unify
proportion	utilize
prospect	valid
publication	violate
purchase	visual
pursue	
recover	
register	
reinforce	

Table 8 shows that there were 108 word families which appeared either in BBC science news or CNN science news.

Figure 2 A sample of academic words in <http://www.bbc.co.uk>.

Artificial skin 'cuts scarring'

A prototype artificial skin used to heal wounds has been developed by British *researchers*.

Writing in the *journal* Regenerative Medicine, UK-based company Intercytex said it had produced promising results in early trials.

It said the skin seemed to *incorporate* itself much better with real tissue than any other skin *substitutes* tried in the past.

The *researchers* hope it might provide an *alternative* to skin grafts.

Currently the best way of treating serious burns and large wounds is to take skin from part of a patient's body and graft it on to the damaged area.

But this is not ideal, and there have been attempts to *create* a form of artificial skin.

However, some doctors say that the failure of these to fully *integrate* with the wound have rendered these efforts of limited value.

Intercytex believes its latest *version* weaves into wounds much better.

1. ICX-SKN is produced from fibrin gel, a blood clotting protein and fibroblast cells, found in human skin. The skin is *created* from a matrix made up of fibrin, a protein found in healing wounds.

To this is added human fibroblasts - cells used by the body to synthesise new tissue.

In a *process* that effectively replicates the way the body makes new skin, the cells produce and *release* another protein, collagen, which makes the matrix more stable.

It is in this form that the "skin" is implanted into a wound.

The *researchers* say that because the matrix is in a *stable* form, it is more able to withstand changes that take place during the healing *process*.

The fact that the collagen is synthesised directly by the cells themselves also more closely mirrors the natural healing *process*.

Quick healing

In tests *researchers* cut an oval section of skin from the arms of six healthy *volunteers* and replaced it with their lab-grown skin.

After 28 days the artificial skin had remained *stable* and the wounds had healed with relatively little scarring.

Dr Paul Kemp, Intercytex's chief scientist, said: "I was very surprised at how quickly the wounds healed.

"If this continues in larger trials then it could *revolutionise* the way in which wounds and burns are treated in the future."

Dr Kemp has been working with Ken Dunn, a consultant surgeon at the burns unit at Wythenshawe Hospital in Manchester.

Mr Dunn said: "This particular product behaves like the patients' own skin.

"It seems to excite much less reaction than the other materials we are using at the moment.

"If this is borne out in larger clinical trials then we would be very interested in using it with our patient group."

Others, however, have warned it is easy to heal a small, surgically-*created* wound in health *volunteers*, and that the true test will come when the *technique* is tried on real patients with real burns.

Dr Phil Stephens, an *expert* in cell biology at Cardiff University, said: "Future studies are needed to *establish* whether this system is substantially better than those already on the market.

"But this skin replacement system has the *potential* to *dramatically* reduce scarring and help heal chronic wounds in aged patients to give them a better quality of life."

Published on 10 July, 2006

The above example shows academic words in <http://www.bbc.co.uk>. There are eighteen academic words (in bold and italics); namely *researchers*, *journal*, *incorporate*, *substitutes*, *alternative*, *create*, *integrate*, *version*, *process*, *release*, *stable*, *revolutionise*, *volunteers*, *technique*, *expert*, *establish*, *potential*, and *dramatically*.

Figure 3 A sample of academic words in <http://www.cnn.com>.

Experts: Global warming behind 2005 hurricanes

MONTEREY, California (Reuters) -- The record Atlantic hurricane season last year can be *attributed* to global warming, several top *experts*, including a leading U.S. government storm *researcher*, said on Monday.

"The hurricanes we are seeing are indeed a direct result of climate change and it's no longer something we'll see in the future, it's happening now," said Greg Holland, a division director at the National Center for Atmospheric Research in Boulder, Colorado.

Holland told a packed hall at the American Meteorological Society's 27th *Conference* on Hurricanes and Tropical Meteorology that the wind and warmer water conditions that fuel storms that form in the Caribbean are "increasingly due to greenhouse gases. There seems to be no other *conclusion* you can *logically* draw."

His *conclusion* will be *debated* throughout the week-long *conference*, as other *researchers* present opposing papers that say changing wind and temperature conditions in the tropics are due to natural events, not the *accumulation* of carbon dioxide emissions clouding the Earth.

Many of the *experts* gathered in the coastal city of Monterey, California, are *federal* employees. The Bush *administration* contends *global* warming is an unproven *theory*.

While many of the *conference's* 500 scientists seem to agree that a warming *trend* in the tropics is causing more and stronger hurricanes than usual, not all agree that *global* warming is to blame.

Some, like William Gray, a veteran hurricane *researcher* at Colorado State University in Fort Collins, Colorado, *attributed* the warming to natural *cycles*.

Gray said he believes salinity buildups and movements with ocean currents cause warming and cooling *cycles*. He *predicted* the Caribbean water will continue to warm for another five to 10 years, then start cooling.

More warming to come

Whatever the cause, computer *projections indicate* the warming to date -- about one degree Fahrenheit (half a degree Celsius) in tropical water -- is "the tip of the iceberg" and the water will warm three to four times as much in the next century, said Thomas Knutson, explaining projections from the National Oceanic and Atmospheric *Administration's* Geophysical Fluid *Dynamics* Laboratory in Princeton, New Jersey.

Adam Lea, a postdoctoral student at Britain's University College London in Dorking, Surrey, presented *research* based on British, German, Russian and Canadian studies that *concludes* half of the increased hurricane activity in the tropics could be *attributed* to *global* warming.

Holland, director of the Mesoscale and Microscale Meteorology Division of the *federal research* center, said tropical storm anomalies in the 1940s and 1950s can be explained by natural variability.

But he said carbon dioxide started changing *traceable* patterns in the 1970s and by the early 1990s, the atmospheric results were *affecting* the storm numbers and *intensities*.

"What we're seeing right now in *global* climate temperature is a signature of climate change," said Holland, a native of Australia. "The large bulk of the scientific *community* say what we are seeing now is linked directly to greenhouse gases."

Hurricane Katrina, which tore onto the Louisiana and Mississippi coasts on August 29, was the deadliest Atlantic hurricane in 77 years and the costliest ever, with property damages *estimated* at \$75 billion.

This year, the weather service's Tropical **Prediction** Center expects more hurricanes than usual, but not as many as last year's record 14.

Published on April 26, 2006.

The above example shows academic words in <http://www.cnn.com>. There are twenty-seven academic words (in bold and italics); namely *attributed*, *attribute*, *experts*, *researcher*, *research*, *conference*, *conclusion*, *conclude*, *logically*, *debate*, *accumulation*, *federal*, *administration*, *global*, *theory*, *trend*, *cycles*, *predicted*, *projections*, *indicate*, *dynamics*, *traceable*, *affecting*, *intensities*, *community*, *estimated* and *prediction*.

In conclusion, the AWL coverage was 7 % of the corpus of BBC science news whereas the academic words in the corpus of CNN science news covered 6 % of the corpus. The AWL coverage of the corpus of BBC science news was one percent higher than that of CNN science news. Four hundred and twenty-two word families occurred in both BBC science news and CNN science news. There were 108 word families which appeared either in BBC science news or CNN science news. Therefore, the academic words in the corpora of BBC science news and CNN science news had some differences. This chapter reported the 50 most frequency word families from the corpora of BBC and CNN science news. The next chapter presents a discussion, interpretation of the results and some suggestions for applications and further studies.

Chapter 5

Conclusion and Discussion

This chapter presents a discussion and interpretation of the results and the implication of the study.

The science news from <http://www.bbc.co.uk> and <http://www.cnn.com> was analyzed using the concordance program to find out the words in the AWL. The words were analyzed to present the frequency of the words in the AWL of the corpora. Then the 50 most frequent academic words were presented. Finally, the words in the AWL in <http://www.bbc.co.uk> and <http://www.cnn.com> were listed and compared.

5.1 Conclusion

Research Question 1: What is the frequency of the words in the AWL in science news?

According to the corpus of BBC science news, there were 6,102 academic words tokens out of the total 89,920 word tokens in the corpus (7 % of the text). The corpus consisted of 476 academic word families out of 570 Coxhead's academic word families (84 % of the total of 570 word families).

In the corpus of CNN science news, there were 6,314 academic words tokens out of the total 110,119 word tokens in the corpus (6 % of the text). The corpus consisted of 476 academic word families out of 570 Coxhead's academic word families (84 % of the total of 570 word families).

Research Question 2: What are the 50 most frequent academic words in science news?

In the corpus of BBC science news, the word *research* had the highest frequency of occurrence (256 word occurrences or 4.20 % of text coverage). The word *source* had the lowest frequency of occurrence (256 word occurrences or 0.46 % of text coverage). The 10 highest frequency of occurrence were the words *research*, *found*, *team*, *energy*, *environment*, *area*, *differentiate*, *data*, *process* and *project* respectively. The 10 lowest frequency of occurrence were the words *source*, *potential*, *involve*, *equip*, *considerable*, *benefit*, *display*, *conclude*, *survey*, *reveal*, *monitor* and *physical* respectively.

In the corpus of CNN science news, the word *research* was the highest frequency of occurrence (216 word occurrences or 3.42% of text coverage). The word *panel* had the lowest frequency of occurrence (31 word occurrences or 0.49% of text coverage). The 10 highest frequency of occurrence were the words *research*, *percent*, *found*, *technology*, *environment*, *create*, *nuclear*, *energy*, *project*, and *team* respectively. The 10 lowest frequency of occurrence were the words *panel*, *normal*, *journal*, *generate*, *equip*, *construct*, *medical*, *fund*, *focus*, *physical* and *final* respectively.

Research Question 3: Are there any differences of the words in the AWL in science news between <http://www.bbc.co.uk> and <http://www.cnn.com> and if they are what are they?

There were some differences of the words in the AWL in science news between <http://www.bbc.co.uk> and <http://www.cnn.com>. The AWL coverage of the corpus of BBC science news was one percent more than the AWL coverage of the corpus of CNN science news. However, both corpora consisted of 476 academic word families out of 570 Coxhead's academic word families (84 % of the total 570 word families). There were 422 word families from Table 7 which appeared in both corpora such as *analyze*, *chemical*, *data*, *environment*, *issue*, *method*, *research* and *theory*. One hundred and eight word

families appeared either in the corpus of BBC science news and CNN science news such as *accommodate, deduce, feature, integrate, proportion, retain, specify, and utilize*.

5.2 Discussion

The results of the research were presented, interpreted and discussed in some detail following the research questions. The results of the study clearly showed that the words in the AWL appeared in the corpora of the BBC science news and CNN science news. The words in the corpus of BBC science news covered approximately 7 % of the text whereas the AWL coverage of CNN science news corpus covered 6 % of the text. These results did not meet Coxhead's 10% coverage of academic texts; nevertheless, these results were close to Coxhead's findings in science discipline. That is, the AWL covered 9.1% of the science subcorpus. Coxhead (2000) claimed that academic words were useful for science students. The number of academic word families in BBC science news corpus was equivalent to CNN science news corpus. Both corpora had 476 academic word families out of 570 word families. The academic words of two corpora accounted for 84 % of Coxhead's academic word families.

The 50 most frequently by used academic words were displayed in this finding. The word *research* was ranked the highest frequency in the two corpora. It was similar to Coxhead's findings. Coxhead divided the academic words into 10 sublists. Sublist 1 comprised the most frequent words in the list whereas sublist 10 comprised the least frequent such as *research, environment, area, data, process, percent* and *create* in sublist 1. These words were also top 10 most frequent words in these findings. In contrast, the frequency of academic words in this research was different from Coxhead's academic word sublists. For instance, the word *team* and *found* were the top

ten most frequent words in this research, but these words were in sublist 9. The findings were probably stated that different corpus had different academic words.

In short, there were some differences between the corpus of BBC science news and CNN science news. The AWL's coverage in CNN science news corpus was only one percent less than BBC science news corpus. There were 422 academic word families appeared in both corpora. There were 108 academic word families that appeared either in the corpus of BBC science news or CNN science news.

5.3 Applications of the study

1. The words in the AWL related to the science field of study can be useful and can be applied in language teaching. Teachers can focus on the frequent academic words to teach the students.

2. Science news can be used as a teaching material in class to enhance students' ability in reading. Since newspaper online relates every day events. Furthermore, Thai students are accustomed to reading news from *the Nation* or *Bangkok post* because they are popular among Thai students. Learning news from the Internet can be another choice for them to improve their reading ability which is very interesting for them.

3. To teach vocabulary, teachers should teach word families to students because students can easily understand or guess the context meaning in various kinds of text.

4. Teachers can apply the concordance to language teaching. Teachers should have students notice word frequency and compare different usages of a word or find phrases and idioms in the texts.

5.4 Limitations of the Study

1. This study focused solely on the words in the AWL in the corpus of BBC science and CNN science news. The AWL's coverage of the text might not provide vocabulary for a higher level in the science discipline.

2. The size of the corpus in this study was small because the corpus contained only 160 news. Therefore, the results may not be generalized.

5.5 Suggestions for Further Studies

1. There should be an analysis in other kinds of text such as health, and law. The results of each text can be compared to find the word in the AWL.

2. Further studies should not focus on only academic words but also the high frequency words and the general service list.

REFERENCES

REFERENCES

- Biber, D., & Conrad, S. (2001,summer). Quantitative corpus-based research: much more than been counting. *TESOL Quarterly*, 35(2):331-340.
- Carter R., & McCarthy, M. (1998). *Vocabulary and Language teaching*. New York: Longman.
- Coady, J. (2002) . L2 vocabulary acquisition through extensive reading. In J. Coady, & T. Huckin (Eds). *Second Language Vocabulary Acquisition* pp.225-237. Cambridge: Cambridge University Press.
- Coxhead, A. (2000,summer). A new academic word list. *TESOL Quarterly*, 34(2):213- 238.
- (2004). An academic word list. Retrieved June 9, 2007, from <http://www.vuw.ac.nz/lals/research/awlinfo.html>.
- Coxhead A., & Nation, P. (2001). The specialized vocabulary of English for academic purpose. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* pp. 252 – 267 Cambridge: Cambridge University Press.
- Donley, M.K., & Reppen, R. (2001,autumn). Using corpus tools to highlight academic vocabulary in SCLT. *TESOL Quarterly*, 10(2/3):7-12.
- Goulden, R., Nation, P & Read, J (1990). How large can a receptive vocabulary be?. *Applied Linguistics* pp. 358-359. Oxford: Oxford University Press.
- Grabe, W. & Stroller, F.L. (2002). Reading and vocabulary development in a second language: A case study. In J.Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* pp.98-112. Cambridge: Cambridge University Press.

- Kaewphanngam, C. (2002). *A corpus analysis of psychology texts as a basis for the development of teaching materials in English Academic Purposes*. Master's thesis, M.A. (Applied Linguistics) Bangkok: Graduate School, Mahidol University.
Photocopied
- Khongphan, W. (2004). *Vocabulary occurrence: An analysis of reading passages in Practice in comprehension II*. Master's project, M.A. (English): Graduate School, Srinakharinwirot University. Photocopied
- Lin, L.H.F. (2002). Overuse, underuse, and misuse: Using concordancing to analyze the use of "it" in the writing of Chinese learners of English. In M. Tan (Ed.), *Corpus studies in language education* pp.63-80 Bangkok: IELE Press.
- Nakprakhon, B. (2005). *Academic vocabulary in political reports: A corpus-based study*. Master's project, Bangkok: Graduate School, Srinakharinwirot University.
Photocopied
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, P. (1994). *New Ways in Tesol Series*. Illinois, USA.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Poonpon, K. (2002) *Vocabulary input in English for science courses: A corpus analysis of intensive and extensive course materials*. Master's thesis, M.A.(Applied Linguistics) Bangkok: Graduate School, Mahidol University. Photocopied
- Sabo, I. (1997) *Approach to vocabulary learning and their relationship to success*. Master's thesis, M.A. Canada: Graduate School, Concordia University.
- Santos, G. (2003) *Three studies on the academic vocabulary knowledge of U.S. language-minority community college students*. Doctor's thesis. D.R. (Education): Graduate School, Harvard University.

- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2002). Using corpora to teach and assess vocabulary. In M. Tan (Ed.), *Corpus studies in language education* pp.31-44 Bangkok: IELE Press.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Tan, M. (2002). *Corpus studies in language education*. Bangkok: IELE Press.
- West, M. (1953). *A general service list of English words*. London: Longman.

APPENDIX

Headwords of the Word Families in the Academic Word List

Numbers indicate the sublist of the Academic Word List (e.g., abandon and its family members are in Sublist 8). Sublist 1 contains the most frequent words in the list, and Sublist 10 contains the least frequent.

abandon	8	alter	5
abstract	6	alternative	3
academy	5	ambiguous	2
access	4	amend	5
accommodate	9	analogy	9
accompany	8	analyse	1
accumulate	8	annual	4
accurate	6	anticipate	9
achieve	2	apparent	4
acknowledge	6	append	8
acquire	2	appreciate	8
adapt	7	approach	1
adequate	4	appropriate	2
adjacent	10	approximate	4
adjust	5	arbitrary	8
administrate	2	area	1
adult	7	aspect	2
advocate	7	assemble	10
affect	2	assess	1
aggregate	6	assign	6
aid	7	assist	2
albeit	10	assume	1
allocate	6	assure	9

attach	6	clarify	8
attain	9	classic	7
attitude	4	clause	5
attribute	4	code	4
author	6	coherent	9
authority	1	coincide	9
automate	8	collapse	10
available	1	colleague	10
aware	5	commence	9
behalf	9	comment	3
benefit	1	commission	2
bias	8	commit	4
bond	6	commodity	8
brief	6	communicate	4
bulk	9	community	2
capable	6	compatible	9
capacity	5	compensate	3
category	2	compile	10
cease	9	complement	8
challenge	5	complex	2
channel	7	component	3
chapter	2	compound	5
chart	8	comprehensive	7
chemical	7	comprise	7
circumstance	3	compute	2
cite	6	conceive	10
civil	4	concentrate	4

concept	1	controversy	9
conclude	2	convene	3
concurrent	9	convert	7
conduct	2	convince	10
confer	4	cooperate	6
confine	9	coordinate	3
confirm	7	core	3
conflict	5	corporate	3
conform	8	correspond	3
consent	3	couple	7
consequent	2	create	1
considerable	3	credit	2
consist	1	criteria	3
constant	3	crucial	8
constitute	1	culture	2
constrain	3	currency	8
construct	2	cycle	4
consult	5	data	1
consume	2	debate	4
contact	5	decade	7
contemporary	8	decline	5
context	1	deduce	3
contract	1	define	1
contradict	8	definite	7
contrary	7	demonstrate	3
contrast	4	denote	8
contribute	3	deny	7
depress	10	dynamic	7

derive	1	economy	1
design	2	edit	6
despite	4	element	2
detect	8	eliminate	7
deviate	8	emerge	4
device	9	emphasis	3
devote	9	empirical	7
differentiate	7	enable	5
dimension	4	encounter	10
diminish	9	energy	5
discrete	5	enforce	5
discriminate	6	enhance	6
displace	8	enormous	10
display	6	ensure	3
dispose	7	entity	5
distinct	2	environment	1
distort	9	equate	2
distribute	1	equip	7
diverse	6	equivalent	5
document	3	erode	9
domain	6	error	4
domestic	4	establish	1
dominate	3	estate	6
draft	5	estimate	1
drama	8	ethic	9
duration	9	ethnic	4

evaluate	2	format	9
eventual	8	formula	1
evident	1	forthcoming	10
evolve	5	foundation	7
exceed	6	found	9
exclude	3	framework	3
exhibit	8	function	1
expand	5	fund	3
expert	6	fundamental	5
explicit	6	furthermore	6
exploit	8	gender	6
export	1	generate	5
expose	5	generation	5
external	5	globe	7
extract	7	goal	4
facilitate	5	grade	7
factor	1	grant	4
feature	2	guarantee	7
federal	6	guideline	8
fee	6	hence	4
file	7	hierarchy	7
final	2	highlight	8
finance	1	hypothesis	4
finite	7	identical	7
flexible	6	identify	1
fluctuate	8	ideology	7
focus	2	ignorance	6

illustrate	3	input	6
image	5	insert	7
immigrant	3	insight	9
impact	2	inspect	8
implement	4	instance	3
implicate	4	institute	2
implicit	8	instruct	6
imply	3	integral	9
impose	4	integrate	4
incentive	6	integrity	10
incidence	6	intelligence	6
incline	10	intense	8
income	1	interact	3
incorporate	6	intermediate	9
index	6	internal	4
indicate	1	interpret	1
individual	1	interval	6
induce	8	intervene	7
inevitable	8	intrinsic	10
infer	7	invest	2
infrastructure	8	investigate	4
inherent	9	invoke	10
inhibit	6	involve	1
initial	3	isolate	7
initiate	6	issue	1
injure	2	item	2
innovate	7	job	4
journal	2	mental	5

justify	3	method	1
label	4	migrate	6
labour	1	military	9
layer	3	minimal	9
lecture	6	minimise	8
legal	1	minimum	6
legislate	1	ministry	6
levy	10	minor	3
liberal	5	mode	7
licence	5	modify	5
likewise	10	monitor	5
link	3	motive	6
locate	3	mutual	9
logic	5	negate	3
maintain	2	network	5
major	1	neutral	6
manipulate	8	nevertheless	6
manual	9	nonetheless	10
margin	5	norm	9
mature	9	normal	2
maximise	3	notion	5
mechanism	4	notwithstanding	10
media	7	nuclear	8
mediate	9	objective	5
medical	5	obtain	2
medium	9	obvious	4

occupy	4	philosophy	3
occur	1	physical	3
odd	10	plus	8
offset	8	policy	1
ongoing	10	portion	9
option	4	pose	10
orient	5	positive	2
outcome	3	potential	2
output	4	practitioner	8
overall	4	precede	6
overlap	9	precise	5
overseas	6	predict	4
panel	10	predominant	8
paradigm	7	preliminary	9
paragraph	8	presume	6
parallel	4	previous	2
parameter	4	primary	2
participate	2	prime	5
partner	3	principal	4
passive	9	principle	1
perceive	2	prior	4
percent	1	priority	7
period	1	proceed	1
persist	10	process	1
perspective	5	professional	4
phase	4	prohibit	7
phenomenon	7	project	4

promote	4	relevant	2
proportion	3	reluctance	10
prospect	8	rely	3
protocol	9	remove	3
psychology	5	require	1
publication	7	research	1
publish	3	reside	2
purchase	2	resolve	4
pursue	5	resource	2
qualitative	9	respond	1
quote	7	restore	8
radical	8	restrain	9
random	8	restrict	2
range	2	retain	4
ratio	5	reveal	6
rational	6	revenue	5
react	3	reverse	7
recover	6	revise	8
refine	9	revolution	9
regime	4	rigid	9
region	2	role	1
register	3	route	9
regulate	2	scenario	9
reinforce	8	schedule	8
reject	5	scheme	3
relax	9	scope	6
release	7	section	1
sector	1	submit	7

secure	2	subordinate	9
seek	2	subsequent	4
select	2	subsidy	6
sequence	3	substitute	5
series	4	successor	7
sex	3	sufficient	3
shift	3	sum	4
significant	1	summary	4
similar	1	supplement	9
simulate	7	survey	2
site	2	survive	7
so-called	10	suspend	9
sole	7	sustain	5
somewhat	7	symbol	5
source	1	tape	6
specific	1	target	5
specify	3	task	3
sphere	9	team	9
stable	5	technical	3
statistic	4	technique	3
status	4	technology	3
straightforward	10	temporary	9
strategy	2	tense	8
stress	4	terminate	8
structure	1	text	2
style	5	theme	8

theory	1	unique	7
thereby	8	utilise	6
thesis	7	valid	3
topic	7	vary	1
trace	6	vehicle	8
tradition	2	version	5
transfer	2	via	8
transform	6	violate	9
transit	5	virtual	8
transmit	7	visible	7
transport	6	vision	9
trend	5	visual	8
trigger	9	volume	3
ultimate	7	voluntary	7
undergo	10	welfare	5
underline	6	whereas	5
undertake	4	whereby	10
uniform	8	widespread	8
unify	9		

VITAE

VITAE

Name:	Miss Luckana Panbanlame
Date of Birth:	May 3, 1979
Place of Birth:	Nakhonpathom
Address:	44/1 M. 9, Tambol Nakhonpathom, Muang Nakhonpathom 73000
Present position:	English teachear Samutsakhonburana School,
Educational Background:	
1996	Mathayom 6 (Mathemetics and Science) The Demonstration of Rajabhat Nakhonpathom Institute School, Nakhonpathom
2001	Bachelor of Education (English) Silpakorn University, Nakhonpathom
2007	Master of Arts (English) Srinakharinwirot University, Bangkok