AN ANALYSIS OF ACADEMIC VOCABULARY IN SPORTS NEWS:

A CORPUS-BASED STUDY

A MASTER'S PROJECT

BY

MALIN DEJTISAK

presented in partial fulfillment of the requirements

for the Master of Arts degree in English

at Srinakharinwirot University

October 2006

Copyright 2006 by Srinakharinwirot University

AN ANALYSIS OF ACADEMIC VOCABULARY IN SPORTS NEWS:

A CORPUS-BASED STUDY

AN ABSTRACT

BY

MALIN DEJTISAK

presented in partial fulfillment of the requirements

for the Master of Arts degree in English

at Srinakharinwirot University

October 2006

Malin Dejtisak. (2006). *An analysis of academic vocabulary in sports news.*

Master's Project, M.A. (English). Bangkok: Graduate School,

Srinakharinwirot University. Advisor Committee: Asst. Prof. Dr. Tipa

Thep-Ackrapong, Dr. Nitaya Suksaeresup, Mr. Stephen Palmer.

This study addressed two questions regarding academic words in the sports

news. It firstly asked about the rate of *Academic Words List* (AWL) which appeared

in sports news collected from *The Nation*, and secondly, what the 50 most frequently

used academic words were in the sports news. The framework of this study was

based on Coxhead's (2000) 570-*Academic Word List* (AWL). A corpus of 102,050

running words taken from a local online newspaper *The Nation* was analyzed against

the AWL of 570 words by using the *Concordance* Program. The results revealed that

words in the AWL covered 1.79% of the total words in the corpus, and the 50 most

frequently used words in the corpus clearly reflected the field of sports news. The

results indicated that sports news might be a good source for English for Specific

Purposes (ESP) learners and provide good text material for Physical Education

students.

การวิเคราะห์คำศัพท์วิชาการในข่าวกีฬาโดยวิธี CORPUS-BASED

บทคัดย่อ
โดย
มาลิน เดชติศักดิ์

มาลิน เดชติศักดิ์ (2549). *การวิเคราะห์คำศัพท์วิชาการในข่าวกีฬาโดยวิธี CORPUS-BASED.*
สารนิพนธ์ ศศ.ม. (ภาษาอังกฤษ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
คณะกรรมการควบคุม: ผู้ช่วยศาสตราจารย์ ดร. ทิพา เทพอัครพงศ์, อาจารย์ ดร. นิตยา
สุขเสรีทรัพย์, มิสเตอร์สตีเฟน พาลเมอร์.

การศึกษาครั้งนี้มีวัตถุประสงค์ที่เกี่ยวเนื่องกับคำศัพท์วิชาการ 2 หัวข้อ ได้แก่ การสำรวจ

จำนวนคำศัพท์วิชาการในข่าวกีฬา จากหนังสือพิมพ์ออนไลน์ เดอะเนชั่น และ ศัพท์วิชาการใดที่

พบมาก 50 อันดับแรก ในการศึกษานี้ ศัพท์วิชาการที่นำมาเป็นกรอบการศึกษาครั้งนี้คือกลุ่มคำ

ศัพท์วิชาการ (Academic Word List) จำนวน 570 คำ ซึ่งพัฒนาโดยค็อกซ์ เฮด(2000)

การศึกษานี้ได้รวบรวมข้อมูลจากข่าวอิเล็กทรอนิคส์เพื่อสร้าง Corpus จำนวน 102,050 คำ และ

วิเคราะห์จำนวนคำศัพท์ดังกล่าวด้วยโปรแกรมคอมพิวเตอร์ Concordance ผลการศึกษาพบ

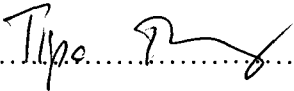คำศัพท์วิชาการ 1.79 เปอร์เซ็นต์ ในข่าวกีฬา และคำศัพท์ที่พบมาก 50 อันดับแรกนั้นส่วนใหญ่

เกี่ยวข้องกับวงการกีฬา ผลการศึกษาชี้ว่าข่าวกีฬาสามารถเป็นแหล่งศึกษาคำศัพท์สำหรับผู้เรียน

ภาษาอังกฤษเฉพาะทาง และสามารถเป็นเอกสารการสอนอ่านแก่นักเรียน นักศึกษาด้านวิชา
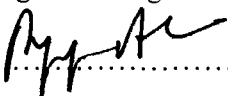
พลศึกษา

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.
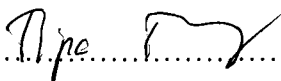

Master's Project Advisor

....

(Assistant Professor Dr. Tipa Thep-Ackrapong)

Chair of the Master of Arts degree in English

....
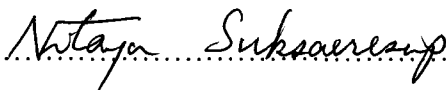
(Dr. Prapaipan Aimchoo)
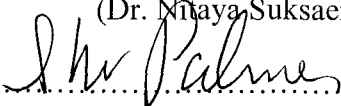
Oral Defense Committee

.... Chair

(Assistant Professor Dr. Tipa Thep-Ackrapong)

.... Reader

(Dr. Nitaya Suksaeresup)

.... Reader

(Mr. Stephen Palmer)

The Graduate School has approved this master's project as partial fulfillment

of the requirements for the Master of Arts degree in English of Srinakharinwirot

University.

.... Dean of the Faculty of Humanities

(Assistant Professor Chaleosri Pibulchol)

October, 13 2006

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

Table   **Page**

LIST OF FIGURE

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Background

In order to be successful in learning vocabulary, teachers and learners should be able to realize which words are necessary to learn. Vocabulary is the knowledge of words and word meanings both in oral and printed language and in productive and receptive forms. Productive form relates to the linguistic skills of speaking and writing, while receptive form relates to the skills of listening and reading. Vocabulary refers to the kind of words language learners must know to comprehend a reading text. Nation (2001) points out that high-frequency used words are the most important in teaching and learning vocabulary, so teachers and learners should recognize them. In learning, learners should know what a word is, what the word sounds like, how the word is pronounced, what the word looks like, and how the word is written and spelled. They should know grammatical patterns, collocations, functions, and meanings of the words. It can be said that second language learners, who learn language for understanding and using it, should study as many words as they can.

Additionally, knowing a word includes being able to recall its meaning when we find it. It also includes being able to see which shade of meaning is most suitable for the context in which it occurs. In addition, knowing the meaning of a word may include making various associations with other related words.

Similarly, in English for Academic Purposes (EAP) programs, knowing which words are worth focusing on during valuable class and independent study time is a challenging matter (Coxhead. 2000). Learners of English for Specific Purposes have

problems understanding textbooks probably because they lack understanding of vocabulary.

Many researchers have attempted to create a word list to encourage language learners to learn words by gathering word families with high frequency of using, ease of learning, coverage of useful concepts, and stylistic level. However, the word list lacked consistency of principles and had many weaknesses. Coxhead developed a new academic word list (henceforth AWL) at the School of Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand. Her list contains 570 word families which have been selected according to several principles. Firstly, in range, the subject areas which the words occur are, for example, Arts, Commerce, Law and Science. Secondly, in frequency, the AWL families have to occur over 100 times in the 3,500,000 word Academic Corpus in order to be considered for inclusion in the list. Thirdly, in uniformity of frequency, the AWL family which occurs a minimum of 10 times in each subject area of the Academic Corpus is included in the list. The AWL has primarily been made for teachers who participate in a program preparing learners for university level study or for students who work alone to learn the words most needed to study at higher education institutions.

The objective of the current research is twofold. First, the AWL developed by Coxhead (2000) will be used to analyze which words are used in sports news. The findings will enable the teacher and learner to select words that should be focused in English for Specific Purposes program. The other objective is academic words. The concordance program will be used to analyze which words in the corpus are academic words.

The Concordance is a computer program that helps language learners and teachers to create concordance lines of any text which has previously been

electronically linked. It is used to make word lists, count word frequency, compare different usages of a word, analyze keywords, and find phrases and idioms. In order to create word lists which occur in a text, the lists are arranged in many different ways. Each word can be optionally accompanied by a count of the number of the time it occurs, and by a note of the percentage of all words which it represents. Concordance is a general-purpose working tool for a close study of text, whether the text is literary, linguistic, historical, religious, philosophical, legal, commercial, or of other kinds. In this study, the concordance program is used to search for academic words in a corpus of sports news.

In short, this study aims to analyze academic vocabulary in a corpus of sports reports by using Coxhead's AWL and concordance program. Also, it creates an option for learners of English for Specific Purposes who want to learn academic vocabulary in sports news.

1.2 Research Questions

1. What is the rate of AWL words that appear in sports news collected from *The Nation*?

2. What are the 50 most frequently used academic words in sports news in the study?

1.3 Scope of the Study

A corpus of 102,050 words was taken from sports news at www.nationmultimedia.com, an on-line database, every Wednesday, Saturday and Sunday during October 1, 2005 to July 31, 2006. The analysis was done in the following aspects:

1. A number of Coxhead's AWL words in the corpus of sports reports was analyzed.

2. The top 50 academic words in the corpus of sports news were studied.

1.4 Significance of the Study

The purpose of this research was to examine the number of academic words in sports news in *The Nation*, a local newspaper in Thailand. The benefits of this study are as follows:

1. It will reveal a source of academic vocabulary.

2. It will create an alternative way for learners of English for Specific Purposes.

1.5 Definition of Terms

1. Academic Words

Academic words are words that are common in different kinds of academic texts, such as *foundation, manipulate, transmit,* and *widespread.*

Coxhead's (2000) 570-*Academic Word List* (AWL), which was compiled from a corpus of 3.5 million running words of written academic text, will be used as an analytical framework of this study.

2. Word Tokens

Word tokens refer to any word forms that occur in the text. That is, individual words occurring more than once in a text are counted each time they are used as word tokens.

3. Sports News

Sports news refers to an online database on sports which is available at www.nationmultimedia.com. The web page provides a direct link to the sports archives which are kept in a text format. The reports are mostly written by Thai people, and are concerned with both national and international sports.

4. Corpus

A corpus refers to a collection of texts which can be in the form of a written text or recorded speech. In this study, a corpus will be created by taking sports news from www.nationmultimedia.com. It comprises 102,050 words of sports reports.

5. Concordance Program

The Concordance is a computer program which is used for analyzing language, making word lists, counting word frequencies, comparing different usages of a word, analyzing keywords, and finding phrases and idioms.

1.6 Overview of the Study

This study aimed at analyzing academic vocabulary in the corpus of sports reports in the following manner. Chapter 2 discusses the literature of vocabulary learning and related research. Then, it presents Coxhead's (2000) Academic Word list. Finally, it discusses studies based on the corpus-based study. Chapter 3 presents the analytical tool and procedure. The analytical tool section discusses the analytical framework: the *Academic Word List* (AWL), and the *Concordance* program. The procedure section deals with the description of the method used in selecting the data and the corpus creation. The procedure employed to analyze the data is also included in this section. Chapter 4 offers the results of the study. Chapter 5 discusses the

results of the analysis in detail and shows how the academic words found in the

corpus of sports reports differs from Coxhead's AWL. It then summarizes the results

by discussing pedagogical implications of the findings and suggestions for further

studies.

CHAPTER 2

REVIEW OF THE RELATED LITERATURE

This chapter focuses on vocabulary learning, related literature, and academic word list, and related literature on the corpus-based study. First, vocabulary learning is discussed. Second, related research on learning vocabulary in particular is studied. Then, Coxhead's (2000) Academic Word list is presented. Finally, corpus-based study and related topics are included.

2.1 Vocabulary Learning

This section presents a survey on vocabulary learning, and other related literature.

The vocabulary knowledge has the following characteristics: it is incremental, and relative. Also it is related to other knowledge, connotations and subtleties.

Nagy and Scott (2000) identify several dimensions that describe the complexity of what it means to know a word. First, word knowledge is incremental, which means that readers need to have many exposures to a word in different contexts before they know it. Second, word knowledge is multidimensional because many words have multiple meanings (e.g., *sage: a wise person; an herb*) and serve different functions in different sentences, texts, and even conversations. Third, word knowledge is interrelated in that knowledge of one word (e.g., *urban*) connects to knowledge of other words (e.g., *suburban, urbanite, urbane*).

As such, knowing a word is a matter of degree rather than an all-or-nothing proposition (Beck & McKeown. 1991; Nagy & Scott. 2000). The degree of knowing a word is reflected in the precision with which we use a word, how quickly we

understand a word, and how well we understand and use words in different modes

(e.g., receptive, productive) and for different purposes (e.g., formal vs. informal

occasions).

Knowing a word also implies knowing how that word relates to other

knowledge. The more we know about a specific concept, the more words we can

bring to our understanding of that concept. Because we have individual interests and

backgrounds, each of us brings different words to shape that understanding.

Finally, knowing a word means being able to appreciate its connotations and

subtleties. When we know a word, we can use and recognize it in idioms, jokes,

slang, and puns.

To sum up, knowing vocabulary involves knowing its characteristics of being

incremental, relative and interrelated. Words' connotations and subtleties are other

useful attributes of vocabulary for learners to recognize the meaning of vocabulary

and appreciate its use in different types of genre.

In short, this section has introduced the characteristics of vocabulary

knowledge learners should know. Next, other related literature of vocabulary learning

is presented.

2.2 Related Literature of Vocabulary Learning

Vocabulary can be learned by listening, speaking, reading and writing. It

encourages language learners to reach their academic goals, and achieve professional

objectives. There are studies on vocabulary learning which focus on its implication,

learning through communication tasks and research on vocabulary growth.

Learners learning vocabulary just to increase vocabulary without giving

attention to its implication may not be successful. Because getting learners to do

language tasks when their vocabulary is inadequate for the task is a frustrating experience (Nation, 1990).

Newton (2001) explains that communication tasks can be productive for vocabulary expansion, either as one-off activities or as rehearsal space for content from other parts of the curriculum. Vocabulary learning can be beneficial for language use and help students pursue communicative goals.

Shostak (2002) points out that vocabulary growth is needed for successful reading comprehension. Therefore, learners need to learn a specific meaning of a word, key information and interesting information about the history of the word, a better understanding of interrelated words in word families, and a better appreciation of language. It also helps learners to understand meanings of a new word within a rich context of supportive and indicative information. Additionally, it gives morphological knowledge, a key strategy that focuses on the structure of a word, and its parts. These are morphemes which include prefixes, suffixes, and roots. Shostak explains that a good understanding of morphology, together with an ability to use definitional and contextual clues, provides a powerful combination that will help learners uncover the precise meaning of words they encounter in direct instruction or in their reading. Solid power of morphology helps students become not only better readers but also better writers. It also helps them to use words sensibly when speaking.

English as a second language learners who are from non-native speaking English countries have difficulty comprehending what they read. A major cause of this difficulty is their lack of understanding new words, especially words that they see in content area textbooks. Cynthia and Johnson (2004) point out that limited vocabulary prevents students from comprehending a text.

Also, the differences of language learners' background affect their language use. Corson (1997) examines the learning and use of academic English words by students who have different social background. He argues that the Graeco-Latin vocabulary of English, which dominates the language's academic vocabulary, offers various levels of potential difficulty for students from different classes, cultures, or linguistic social groups. Corson points out that people from different backgrounds use academic words in different ways; these differences increase during peak periods of word acquisition, such as adolescence. L1 and L2 students who have different exposures to language produce different arrangements of the mental lexicon which can lead to sociocultural variations in the learning and use of words. Saville-Troike (1984) reports that vocabulary knowledge is the most important aspect of oral proficiency for academic achievement in learning another language.

Additionally, words are only fully learned when they are available for active use (Corson. 1997). When words are set only in passive vocabulary which inhabits their active use, full word learning will not take place. Furthermore, difficulties in identifying words may even affect passive vocabulary acquisition.

For instance, many readers when encountering academic words, which mostly are Graeco-Latin, for the first time, often ignore these words. They see them as alien, strange, and forbidding, not just because they are new words used to express ideas within unfamiliar meaning systems, but also because they are not acquired at the same time as their awareness skills needed to enable them to translate and memorize these words in their mind, which would help realize the meanings and functions for the ease of activation and future use.

On the other hand, the awareness of English lexical rules is useful for language learners. Corson (1997) suggests that when people are more aware of the

derivational rules of English, the learning and use of academic Graeco-Latin words become easier. This is a special kind of knowledge about English. He explains that these language awareness skills offer different levels of organization that are supplementary to other ways of analyzing words. These certainly make learning Graeco-Latin words easier. Some researchers have suggested that new words are best understood by analogy to similar known words (Anglin. 1993). Some have proposed rule-based models of analysis: giving meaning to each part of a word and combining them (White, Power, & White. 1989). It can be said that one technique probably suits some words, and other techniques suit some other words.

The other significant point about learning academic vocabulary is the benefits of learning other languages. Corson (1997) states that studying and learning L2s other than English seems very beneficial for academic vocabulary development. For instance, studying Latin and Ancient Greek used to offer vocabulary advantages to students of English. These advantages still seem relevant to learning and using morphologically complex and low-frequency words. However in the present day, these advantages are no more widely available elsewhere, through the study of Romance languages.

In short, vocabulary learning should be based on communication tasks. Moreover, to promote vocabulary growth, the learners should be encouraged to learn word meanings and formation. However, different results in learning words may lie in the learners' different social backgrounds. If words are used actively, they will enhance vocabulary development. Finally awareness of academic lexical rules also fosters vocabulary growth and other aspects of language learning. Next, information of Coxhead's (2000) Academic Word List is presented.

2.3 Coxhead's (2000) Academic Word List

Coxhead's (2000) Academic Word List was developed from the needs for a new academic word list based on data gathered from a large, well designed corpus of academic English. Coxhead improved the AWL by considering words representing the texts of interest, the organization of the corpus, its size, and the criteria used for word selection. The AWL includes 570 word families that comprise a specialized vocabulary with a good coverage of academic texts, regardless of the subject area. It accounts for 10% of the total tokens in the Academic Corpus, and more than 94% of the words in the list occur in 20 or more of the 28 subject areas of the Academic Corpus. In the next part, the corpus-based study and related topics are discussed.

2.4 The Corpus-based Study

A corpus is a collection of linguistic data, either a written text or recorded speech, which can be used as a starting point of linguistic description or as a means of verifying hypotheses about a language (Crystal. 1991). The first language corpora were compiled decades ago, and many corpus-based linguistic studies have been conducted since then (Biber & Conrad. 2001). Some of the earliest uses of corpus linguistics were for applied purposes, especially the compiling of dictionaries. The unifying characteristics of the corpus-based research include the use of a large, representative electronic database of spoken or written texts, or both, and the use of computer-assisted analysis techniques.

Biber (1998) briefly outlines main features of a corpus-based approach which are as follows:

1. It analyses actual patterns of language use in authentic data.

2. It makes use of a large collection of corpus data taken from written and spoken texts or both.

3. It utilizes computer concordancing and tagging programs for the analysis.

4. It relies on corpus linguistic principles of analysis to report findings.

In brief, this section has presented a corpus which is a body or collection of texts. It can be in the form of written text or recorded speech. The corpus is used for analyzing patterns of language use and making a collection of data taken. In order to analyze words in the corpus, the computer concordancing and tagging programs are used.

2.5 Previous Research

As part of previous research, Nakprakhon (2005) has studied the coverage and the top 50 academic words in a corpus of political reports. He analyzed a corpus of 122,811 running words from *The Nation* against the *Academic Word List* (AWL) by using the Concordance program. He found that there were 5.7% academic words in the corpus, and the top 50 words in the AWL were clearly political-related words.

So far, the literature reviewed has discussed vocabulary growth, Coxhead's AWL and previous research. In the first regard, communication tasks, active word use, lexical meaning and formation rules are suggested to promote learners' vocabulary development. In the second regard, Coxhead's AWL has been described. Based on a large authentic linguistic data, the AWL offers a new list of 570 academic word families, which will be used to determine academic words in the present study. Finally one previous corpus-based study has been reviewed. The results showed that

the political news under investigation could be considered academic text because it

contained an acceptable rate of academic words.

The next chapter will present the methodology used in this study.

CHAPTER 3

METHODOLOGY

This chapter consists of two main sections: the analytical tool and procedure. The analytical tool section discusses the analytical frameworks: the *Academic Word List* (AWL), and the *Concordance* program. The procedure section deals with the description of the method used in selecting the data and creating a corpus. This section also includes the procedure employed to analyze the data.

3.1 Analytical Tool

3.1.1 Analytical Framework

The framework of this study is based on Coxhead's (2000) 570-*Academic Word List* (AWL), which was compiled from a corpus of 3.5 million running words of written academic text.

3.1.2 The Concordance Program

The Concordance Program is a computer program which is used for analyzing the language that is used in real situations. It is used to make word lists, count word frequencies, compare different usages of a word, analyze keywords, and find phrases and idioms. This program is specifically used for analyzing academic words in the corpus of sports news in this research.

Figure 1 A Sample of the Concordance Program
Source: http://www.rjcw.freeserve.co.uk/

3.2 Procedure

The research was conducted in accordance with the following steps:

3.2.1 Data Collection

The data used in this study were sports news taken from on-line database at

www.nationmultimedia.com from October 1, 2005 to July 31, 2006. Wednesday,

Saturday and Sunday reports were chosen because there were various types of news

reported on these days. Totally, there were 248 reports.

3.2.2 Corpus Creation

The steps in creating the corpus were as follows:

a) Sports news from the website was copied. The homepage of the website showed links to different archives which were categorized into smaller sections, such as entertainment, sports, education, health, politics, and editorial. Then the sports news texts were copied one at a time.

b) Dates and months in the news were removed from the texts. Only headlines and the body of the news were copied into a master file. For instance, there was a reference on the top of every page saying, "published on (date/month/year)." This phrase would affect the frequency of word list because the word *publish* was one of the words in the academic wordlist.

c) The texts were inserted into the sports news file. Then the corpus was fully developed.

d) The corpus then was transferred into a hard-drive in a personal computer for the analysis.

3.2.3 Data Analysis

A concordance list of the corpus was produced by using the *Concordance* program in order to answer the two main research questions in this study.

Research question 1: What is the rate of AWL words that appear in sports news collected from *The Nation*?

The steps used to analyze the data for answering this question were as follows:

a) The total tokens of the corpus were searched by using the option *Make Full Concordance.*

b) The total word tokens of the corpus were presented to the program.

c) All word tokens the Concordance generated from the corpus were stored.

d) The total AWL words were calculated by using the Concordance's Pick

List tool. This tool added the total words of the AWL, headwords and

their families. Then the AWL word tokens were acquired.

e) The percentage of the AWL was obtained by using the following formula:

$$\text{The percentage of the AWL} = \frac{\text{total AWL word tokens X 100}}{\text{total word tokens in the corpus}}$$

Research question 2: What are the 50 most frequently used academic words in sports news?

The following steps were used for answering this question:

a) The frequency of each word in the AWL acquired from the process in the

first research question was manually compared.

b) The 50 most frequently used words were determined and stored.

c) The 50 most frequently used words were reported.

The methodology of this study was based on two analytical tools. One analytical tool was Coxhead's (2000) *Academic Word List* (AWL) as a framework. Also, the *Concordance* program was used as a tool to analyze a large corpus academic words of over 100,000 words. The percentage of the AWL was obtained by using total AWL words tokens multiplied by 100 and divided by total word tokens in the corpus. The next chapter will present the results of this study.

CHAPTER 4

FINDINGS

This chapter presents the findings of the study which are organized so as to answer the following research questions:

1. What is the rate of academic words that appear in sports news collected from *The Nation*?

2. What are the 50 most frequently used academic words in sports news in the study?

The results of the study are presented and discussed in two parts:

Part 1 presents the broader picture of the corpus of sports news, such as the total word tokens and total word types.

Part 2 presents the tables related to the questions of the study.

In order to explore the number of academic words in the corpus of sports news, the word tokens of both academic words and the entire corpus of sports news were compiled.

Table 1 Academic word types, academic word tokens, total word types and tokens, and the coverage of academic words in the corpus of sports news.

Type of words	Number of words
Academic word types	216
Academic word tokens	1,834
Total word types in the corpus	11,555
Total word tokens in the corpus	102,050
Academic words coverage (%)	1.79%

Table 1 shows academic word types, academic word tokens, total word types and tokens, and the coverage of academic words in the corpus. Academic word types in the corpus accounted for 1.8% of all word types in the corpus (216 out of 11,555), while academic word tokens accounted for 1.79% of all word tokens (1,834 out of 102,050). That is to say, the 1.79% figure was the coverage of academic words in the corpus of sports news. To sum up, the academic words in the corpus were less than Coxhead's (2000) AWL 2% coverage of academic texts, and the number of word types in the corpus which was 11,555, shows vocabulary size in the sports news' corpus.

In order to answer the second question, the 50 most frequently used academic words in sports news, the frequency of academic headwords were compiled and counted as one word as presented in Table 2.

Table 2 The 50 most frequently used academic words in the corpus of sports news

Word families	Frequency	Percent
1. team	394	21.48
2. final	184	10.03
3. overall	85	4.63
4. despite	39	2.12
5. professional	36	1.96
6. series	34	1.85
7. previous	32	1.74
8. category	27	1.47
9. found	26	1.42
10. goal	26	1.42
11. challenge	24	1.30
12. target	23	1.25
13. major	19	1.03
14. authority	18	0.98
15. media	18	0.98
16. couple	17	0.92
17. annual	16	0.87
18. individual	16	0.87
19. style	16	0.87
20. task	15	0.82
21. role	14	0.76
22. so-called	14	0.76
23. contract	13	0.71
24. classic	13	0.71
25. formula	11	0.59
26. job	11	0.59
27. channel	10	0.54
28. feature	10	0.54
29. plus	10	0.54
30. available	9	0.49
31. complex	9	0.49
32. error	9	0.49
33. format	9	0.49
34. partner	9	0.49
35. preliminary	9	0.49
36. range	9	0.49
37. area	8	0.43
38. encounter	8	0.43
39. fund	8	0.43
40. promote	8	0.43
41. status	8	0.43
42. affect	7	0.38

Table 2 (continued)

Word families	Frequency	Percent
43. aggregate	7	0.38
44. crucial	7	0.38
45. focus	7	0.38
46. impact	7	0.38
47. maintain	7	0.38
48. schedule	7	0.38
49. academy	6	0.32
50. compound	6	0.32

Table 2 shows the 50 most frequently used academic words in the corpus of sports news. Some words have the same frequency of use.

From the study, the most frequently used 50 words clearly reflected the field of the corpus of sports news, e.g. *team, final, professional, series* and *goal*. The word *team* had the highest frequency of occurrence (394 or 21.48%), while the words *academy* and *compound* had the lowest frequency of occurrence (6 or 0.32%).

In order to examine academic vocabulary which can be found in a text, an example of sports news in the corpus is presented.

An example of academic vocabulary found in the sports news

A DREAM COMES TRUE

Published on May 3, 2006

Dott champ after an epic thriller

It was glory for Graeme Dott in Sheffield's Crucible Theatre early yesterday when he became the 888.com world champion following an 18-14 victory over Peter Ebdon.

The victory won him a $200,000 pay day.

After more than 13.5 hours of sometimes painstaking snooker, Dott finally scraped home after Ebdon, 15-7 down, pulled back to 15-13.

But the 2002 world champion who is now based in Dubai was never quite able to make up the lost ground, though in the end it was a desperately close call.

"The 68 clearance I made to go 17-14 was probably the best I've ever made. I just don't know where it came from," said Dott, a 50-1 chance to win the title at the start of the championship 17 days ago.

Ebdon who could be backed at 20-1, and Dott had been embroiled in a record-breaking 27th frame.

When Ebdon eventually potted the black to complete a clearance from the green of 25, the frame had taken 74 minutes, 08 seconds to complete, the longest ever in front of the television cameras.

That pushed to one side Steve Davis and Dene O'Kane's 73 minute, 30 seconds *previous* longest TV frame set at the 1989 World *Team* Cup at Bournemouth, while the 69 minutes-long world *final* deciding frame shoot-out between Davis and Dennis Taylor in 1985 paled into history.

But the *final* brought about an amazing transformation as Ebdon, 15-7 behind going into the last set of 13 frames, clawed his way back to 15-13 and right in the thick of things, a break of 84 took him to within two of the Scot for the first time since Dott led 3-1 and 4-2. Eventually Dott stopped Ebdon's gallop by taking frame 29 with a break of 66 to lead 16-13, only two short of the winning post.

Snooker and boxing entrepreneur Barry Hearn had issued a warning to the finalists before play started on the last day not to expect an invitation to next season's Premier League unless standards improved in the chase for the $200,000 first prize.

It was unlikely that either player was *aware* of Hearn's comments or even interested before the *final* resumed with Dott ahead 11-5 overnight and later extending this to eight frames.

Up until then Ebdon, the winner in 2002 and runner-up in 1996, had experienced a nightmare, doing very little right and looking incapable of making a serious *challenge* to Dott's supremacy.

The Scot, winning five of six frames, recorded breaks of 65, 62 and 56 to get within three of what would be his first tournament victory.

Ebdon, though, is a fighter who does not know the meaning of the word defeat and as the *final* moved into the concluding stages, slowly but surely he pushed himself back into the reckoning.

His first *task* was to win frame 23 with the first century of the *final*, a clearance of 117 and with Dott unsure of himself, mistakes followed and a promising break of 51 let Ebdon in to snatch the next frame out of the fire with a clearance to the pink of 32.

Dott by now was getting shakier and shakier and it was no surprise when he also lost the next two frames to go in at the *interval* only four behind at 15-11.

Frame 25 saw Ebdon, 33-29 in front, record a decisive break of 34 and in the 26th when he was 38-4 behind, he returned to the table after Dott missed a red into a top pocket to *compile* a 66 to signal the *final* was very much alive again.

The tempo increased on day two and after 20 frames had been completed, the total running time was eight hours, 16 minutes, 56 seconds and the average time per frame was down from 28 minutes to 25 minutes, where it was to stay.

"If this was a boxing fight, it would have been stopped on Sunday night to spare the fans any more punishment. "They said the first World War would be over by Christmas and on Monday morning they were talking about this match in the same terms," added Hearn.

The above example shows academic words which can be found in sports news. There are eight academic words in the news (in bold and italics), which are; *aware, challenge, compile, final, interval, previous, task* and *team*. These words occur once each; only *final* has occurred six times in the example news.

It can be concluded from the results in Table 2 that the number of academic words in the corpus of sports news was lower than those in Coxhead's (2000) AWL. There were 1,834 academic word tokens or 216 academic word types found in the corpus of sports news, but 354 academic words did not occur in the corpus. The disappearance of these words can be explained that some words may be found frequently in a particular field, for example, in scientific texts (*hypothesis, mechanism, react*), or political reports (*constitute, policy, ministry*). However, this is not surprising because academic words, as the name suggests, are naturally found in academic texts. From the study, there are twelve words which have occurred more

than 20 times. They are *team, final, overall, despite, professional, series, previous, category, found, goal, challenge* and *target*. In addition, the words *team* and *final* have occurred more than 100 times.

The next chapter will discuss the results, some suggestions for applications and further studies.

CHAPTER 5

CONCLUSION AND DISCUSSION

In this chapter, the conclusion of the results of the analysis and how AWL words found in the corpus of sports reports differs from Coxhead's AWL are presented. Then a summary of the results and pedagogical implications of the findings and suggestions for further studies are discussed.

5.1 Conclusion

This section presents answers of the two research questions.

Research question 1: What is the rate of AWL words that appear in sports news collected from *The Nation*?

The AWL words accounted for 1.79% of all word tokens in the corpus of sports news. There were 1,834 academic word tokens out of the number of 102,050 word tokens in the corpus, and most of them occurred in an average of 26.7 times. The AWL coverage in the corpus was about 8.2% less than that of Coxhead's (2000) coverage of Academic Corpus (10%).

Research question 2: What are the 50 most frequently used academic words in sports news in the study?

The 50 most frequently used academic words in the corpus related to the field of sports are *team, final, professional, series, goal* and *challenge*. From the corpus, *team*, the most frequently used academic word, occurred 394 times or accounted for 21.48%. It was followed by *final* (184 frequency or 10.03%) and *overall* (85

frequency or 4.63%). The least frequently used words in the top 50 of the corpus were *accompany* and *compound*, which had the same frequency of use or 0.32%.

5.2 Discussion

As mentioned in Chapter 1, the findings of this study would reveal a source of academic vocabulary and create an alternative way for learners of English for Specific Purposes who may study sports news in particular. In addition, the results offer important characteristics of academic vocabulary in the corpus of sports news and should be taken into account in deciding whether this kind of text should be used as teaching materials for language teachers. According to the study, the number of academic words in sports reports were different from Coxhead's (2000) study. Coxhead claims that her AWL includes 570 word families which account for 10% of the total tokens in the Academic Corpus, and more than 94% of the words in the list occur in 20 or more the 28 subject areas of the Academic Corpus. Nevertheless the academic words from the AWL has only 1.79% coverage of the corpus of sports reports from *The Nation* online news database. The results showed that the rate of academic words was too low, less than 2%. Therefore, it should not be considered an academic text. As a previous study in the corpus of political report, the AWL words accounted for 5.7% (Nakprakhon. 2005). There were 7,020 out of the total 122,811 word tokens in the corpus. The AWL coverage in the present corpus is about 4% less than that of Coxhead's (2000), and 3% less than her specific findings on the coverage of vocabulary in science discipline. Because the corpus of sports news is not a good academic text, it should be used as an ESP text, especially for physical education students.

5.3 Applications

1. The *Concordance* program is very useful for language teaching and learning. Teachers may use concordance lines in class to highlight some words in the AWL in context by saving texts in plain text files and use them to make a concordance form. They can present the learners with appropriate contexts for the use of vocabulary. Moreover, the highlighted vocabulary always comes with collocations that learners can understand in different situations. The program can show in the Context View by clicking on a line; the full text from which the concordance was made will display the position of that word. Also, it can show the full text viewer by double-clicking on any word; the Wordlist View will position itself to display that word.

2. The *Concordance* program is also beneficial for language learners. They can use the interactive concordance for self-study of vocabulary. At the first stage, teachers may explain how to use the program to students so that they will be able to use the program efficiently.

3. The academic vocabulary in the corpus of sports news which has been found in this study may be useful for study units in English for Specific Purposes courses. Teachers may make students familiar with those words by creating language activities, such as reading or listening practices.

In conclusion, this section has provided the value of the *Concordance* program for language teaching and learning. Students can study vocabulary from teachers, or they can have a self-study. Also, the data from this study may be valuable information for teachers and learners in English for Specific Purposes program.

5.4 Limitations of the Study

1. The data collected from this study might be regarded as local-oriented news from only one source, an online-database of English newspaper in Thailand *The Nation*. In addition, most of the news in the corpus was written by local Thai people. Therefore, the corpus might not be representative of sports news as compared with other sources, for example, *CNN* or *BBC*.

2. The results of this study might not be exactly the same as other results from similar research in the future because the data were unique and dependent on the situations occurring only at a certain period of time. For example, in November 2005, there was the SEA Games in Vietnam; and during the months of June and July 2006, there was the World Cup in Germany.

3. According to the small size of the corpus, the findings cannot be easily generalized. A study on a bigger corpus may yield different results.

5.5 Suggestions for Further Studies

1. This study was conducted with only a corpus of sports news. There should be similar research with another genre, such as advertising or business. Also, there should be research with a bigger size of the sports news with different kinds of sources: *CNN, BBC, The Sun*, etc.

2. There should be a comparative study of academic words in a corpus of sports news in Thai local newspaper and international newspaper.

REFERENCES

REFERENCES

Biber, D. & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly, 35*(2): 331-340.

Cobb, T. & Horst, M. (2001). Growing academic vocabulary with a collaborative on-line database. Retrieved April 24, 2006. from http://elc.polyu.edu.hk/conference/papers2001/cobb.htm.

Corson, D. (1997). The learning and use of academic English words. *Language Learning, 47*(4): 671-718.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2): 213-238.

Coxhead, A. (2004). An academic word list. Retrieved April 10, 2006, from http://www.vuw.ac.nz/lals/research/awl/info.html.

Cummins, J. (2000). Academic language learning, transformative pedagogy, and information technology: Towards a critical balance. *TESOL Quarterly, 34*(3): 537-548.

Hyland, K. (2002). Options of identity in academic writing. *ELT Journal, 56*(4): 351-358.

Nakprakhon, B. (2005). *Academic vocabulary in political reports: A corpus-based study*. Unpublished thesis, Srinakharinwirot University, Bangkok, Thailand.

Napoli, D. J. (2003). *Language matters: A guide to everyday questions about language*. Oxford: Oxford University Press.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Massachusetts: Heinle and Heinle Publishers.

Newton, J. (2001). Options for vocabulary learning through communication tasks. *ELT Journal, 55*(1): 30-37.

Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly, 18*(2): 199-219.

Shostak, J. (2004). The value of direct and systematic vocabulary instruction. Retrieved May 31, 2006, from http://www.sadlier-oxford.com/pdf/9147-9_VW_WhitePaper_Vol7.pdf.

Sinclair, S. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sutarsyah, C. (1994). How useful is EAP vocabulary for ESP? A corpus based study.

RELC Journal, 25(2): 34-50.

Tan, M. (2002). *Corpus Studies in Language Education.* Bangkok: IELE Press.

Watt, R. J. C. (2004). Concordance program. Retrieved May 5, 2006, from
http://www.rjcw.freeserve.co.uk/

Wei, Y. (1999). *Teaching collocations for productive vocabulary development.*
Paper presented at the Annual Meeting of the Teachers of English to Speakers
of Other Language, New York.

APPENDI

Coxhead's (2000) Academic Word List

The Coxhead's (2000) 570-Academic Word List (AWL) was used as a framework of this study.

abandon	assess	comment
abstract	assign	commission
academy	assist	commit
access	assume	commodity
accommodate	assure	communicate
accompany	attach	community
accumulate	attain	compatible
accurate	attitude	compensate
achieve	attribute	compile
acknowledge	author	complement
acquire	authority	complex
adapt	automate	component
adequate	available	compound
adjacent	aware	comprehensive
adjust	behalf	comprise
administrate	benefit	compute
adult	bias	conceive
advocate	bond	concentrate
affect	brief	concept
aggregate	bulk	conclude
aid	capable	concurrent
albeit	capacity	conduct
allocate	category	confer
alter	cease	confine
alternative	challenge	confirm
ambiguous	channel	conflict
amend	chapter	conform
analogy	chart	consent
analyse	chemical	consequent
annual	circumstance	considerable
anticipate	cite	consist
apparent	civil	constant
append	clarify	constitute
appreciate	classic	constrain
approach	clause	construct
appropriate	code	consult
approximate	coherent	consume
arbitrary	coincide	contact
area	collapse	contemporary
aspect	colleague	context
assemble	commence	contract
contradict	distort	expose
contrary	distribute	external

contrast	diverse	extract
contribute	document	facilitate
controversy	domain	factor
convene	domestic	feature
converse	dominate	federal
convert	draft	fee
convince	drama	file
cooperate	duration	final
coordinate	dynamic	finance
core	economy	finite
corporate	edit	flexible
correspond	element	fluctuate
couple	eliminate	focus
create	emerge	format
credit	emphasis	formula
criteria	empirical	forthcoming
crucial	enable	foundation
culture	encounter	found
currency	energy	framework
cycle	enforce	function
data	enhance	fund
debate	enormous	fundamental
decade	ensure	furthermore
decline	entity	gender
deduce	environment	generate
define	equate	generation
definite	equip	globe
demonstrate	equivalent	goal
denote	erode	grade
deny	error	grant
depress	establish	guarantee
derive	estate	guideline
design	estimate	hence
despite	ethic	hierarchy
detect	ethnic	highlight
deviate	evaluate	hypothesis
device	eventual	identical
devote	evident	identify
differentiate	evolve	ideology
dimension	exceed	ignorance
diminish	exclude	illustrate
discrete	exhibit	image
discriminate	expand	immigrate
displace	expert	impact
display	explicit	implement
dispose	exploit	implicate
distinct	export	implicit
imply	label	objective
impose	labour	obtain

incentive	layer	obvious
incidence	lecture	occupy
incline	legal	occur
income	legislate	odd
incorporate	levy	offset
index	liberal	ongoing
indicate	licence	option
individual	likewise	orient
induce	link	outcome
inevitable	locate	output
infer	logic	overall
infrastructure	maintain	overlap
inherent	major	overseas
inhibit	manipulate	panel
initial	manual	paradigm
initiate	margin	paragraph
injure	mature	parallel
innovate	maximise	parameter
input	mechanism	participate
insert	media	partner
insight	mediate	passive
inspect	medical	perceive
instance	medium	percent
institute	mental	period
instruct	method	persist
integral	migrate	perspective
integrate	military	phase
integrity	minimal	phenomenon
intelligence	minimise	philosophy
intense	minimum	physical
interact	ministry	plus
intermediate	minor	policy
internal	mode	portion
interpret	modify	pose
interval	monitor	positive
intervene	motive	potential
intrinsic	mutual	practitioner
invest	negate	precede
investigate	network	precise
invoke	neutral	predict
involve	nevertheless	predominant
isolate	nonetheless	preliminary
issue	norm	presume
item	normal	previous
job	notion	primary
journal	notwithstanding	prime
justify	nuclear	principal
principle	reveal	supplement
prior	revenue	survey

priority
proceed
process
professional
prohibit
project
promote
proportion
prospect
protocol
psychology
publication
publish
purchase
pursue
qualitative
quote
radical
random
range
ratio
rational
react
recover
refine
regime
region
register
regulate
reinforce
reject
relax
release
relevant
reluctance
rely
remove
require
research
reside
resolve
resource
respond
restore
restrain
restrict
retain
volume
voluntary

reverse
revise
revolution
rigid
role
route
scenario
schedule
scheme
scope
section
sector
secure
seek
select
sequence
series
sex
shift
significant
similar
simulate
site
so-called
sole
somewhat
source
specific
specify
sphere
stable
statistic
status
straightforward
strategy
stress
structure
style
submit
subordinate
subsequent
subsidy
substitute
successor
sufficient
sum
summary

survive
suspend
sustain
symbol
tape
target
task
team
technical
technique
technology
temporary
tense
terminate
text
theme
theory
thereby
thesis
topic
trace
tradition
transfer
transform
transit
transmit
transport
trend
trigger
ultimate
undergo
underlie
undertake
uniform
unify
unique
utilise
valid
vary
vehicle
version
via
violate
virtual
visible
vision
visual

welfare
whereas
whereby
widespread

VITAE

VITAE

Name:	Miss Malin Dejtisak
Date of Birth:	June 16, 1978
Place of Birth:	Bangrak, Bangkok
Address:	38 Charoenkrung 34 Road, Bangrak, Bangkok 10500
Present Position:	Lecturer
	International College for Sustainability Studies,
	Srinakharinwirot University, 114 Sukhumvit 23 Road,
	Bangkok 10110

Educational Background:

2006	Master of Arts (English)
	Srinakharinwirot University, Bangkok
2001	Bachelor of Arts (Business English)
	Rajabhat Institute Suan Dusit, Bangkok
1996	Mathayom 6 (English and French)
	Santa Cruz Convent School, Bangkok