

ACADEMIC VOCABULARY IN POLITICAL REPORTS:

A CORPUS-BASED STUDY

A MASTER'S PROJECT

BY

MR. BANCHHA NAKPRAKHON

presented in partial fulfillment of the requirements

for the Master of Arts degree in English

at Srinakharinwirot University

March 2005

428.1

12.34

5.12

**ACADEMIC VOCABULARY IN POLITICAL REPORTS:
A CORPUS-BASED STUDY**

AN ABSTRACT

BY

MR. BANCHHA NAKPRAKHON

presented in partial fulfillment of the requirements

for the Master of Arts degree in English

at Srinakharinwirot University

March 2005

Bancha Nakprakhon. (2005). *Academic vocabulary in political reports*. Master's Project, M.A. (English). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Dr. Saengchan Hemchua, Assistant Professor Dr. Tipa Thep-Ackrapong, Mr. Michael Rudy.

This study addressed two questions regarding academic words in the political reports environment. It firstly asked whether words in the *Academic Word List* (AWL) were significantly high in the political reports, and secondly, what the top 50 academic words were in the corpus of political reports. The corpus of 122,811 running words taken from the 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 5.7 % of the total words in the corpus, and the top 50 words in the AWL were clearly political-related words. The corpus contained most of the AWL word families (513 words out of 570 words accounting for 90%). The results indicated that reading political reports might produce positive results in preparing learners for academic vocabulary at the university level.

คัพพิชากการในชาวการเมือง

บทคัดย่อ

ของ

นายบัญชา นาคประโคน

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต สาขาวิชาภาษาอังกฤษ
มีนาคม 2548

บัญชา นาคประโคน (2548). ศัพท์วิชาการในข่าวการเมือง. สารนิพนธ์ ศศ.ม. (ภาษาอังกฤษ).

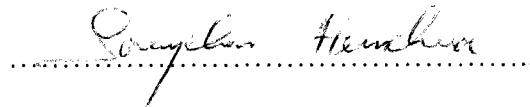
กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม:

ดร. แสงจันทร์ เหมเชื้อ, ผศ.ดร.ทิพา เทพอัศวพงศ์, มิสเตอร์ ไมเคิล รูดี.

การศึกษาคำศัพท์ครั้งนี้มีวัตถุประสงค์เพื่อสำรวจจำนวนคำศัพท์วิชาการในข่าวการเมือง จากหนังสือพิมพ์ออนไลน์ เดอะเนชั่น และศัพท์วิชาการใดที่พบมาก 50 อันดับแรก ศัพท์วิชาการที่นำมาเป็นกรอบครั้งนี้คือกลุ่มคำศัพท์วิชาการ (Academic Word List) 570 คำที่พัฒนาโดย ค็อกซ์เฮด(2000) โดยในการศึกษาคำศัพท์ครั้งนี้ได้รวบรวมข้อมูลซึ่งอยู่ในรูปของข่าวอิเล็กทรอนิกส์เพื่อสร้างเป็นชุดข้อมูล (Corpus) จำนวน 122,811 คำ แล้วนำไปวิเคราะห์หาจำนวนคำศัพท์ โดยใช้โปรแกรมคอมพิวเตอร์ Concordance ผลการศึกษาคำศัพท์วิชาการ 5.7 เปอร์เซ็นต์ ในข่าวการเมือง และ คำศัพท์วิชาการ 50 อันดับแรกนั้น ส่วนใหญ่เป็นคำที่เกี่ยวข้องกับการเมือง นอกจากนี้ ยังพบศัพท์วิชาการ 513 คำ จากจำนวน 570 คำที่ใช้เป็นกรอบในการศึกษา ซึ่งคิดเป็น 90 เปอร์เซ็นต์ คำศัพท์วิชาการ 50 อันดับแรกนั้นนับว่าเป็นคำศัพท์ที่มีประโยชน์ในการเรียนการสอนภาษาอังกฤษ เพราะเป็นคำศัพท์ที่เกี่ยวข้องกับเนื้อหาของข่าว ซึ่งจะช่วยให้ผู้อ่านข่าวการเมืองได้เข้าใจมากขึ้น อีกทั้งยังเป็นคำศัพท์ที่สามารถพบได้ในหนังสือวิชาการทั่วไป ซึ่งถือเป็นกลุ่มคำศัพท์วิชาการพื้นฐานที่มีประโยชน์ในการเรียนการสอนในระดับอุดมศึกษา

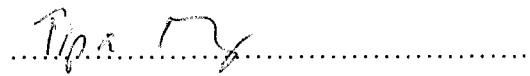
The Master's Project Committee and Oral Defense Committee have approved this master's project as partial fulfillment of the requirements of the Master of Arts degree in English of Srinakharinwirot University.

Master's Project Advisor



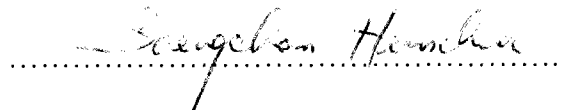
(Dr. Saengchan Hemchua)

Chair of the Master of Arts degree in English



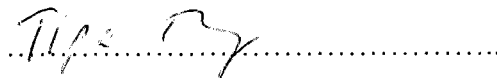
(Assistant Professor Dr. Tipa Thep-Ackrapong)

Oral Defense Committee



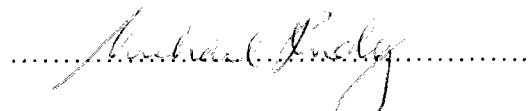
(Dr. Saengchan Hemchua)

Chair



(Assistant Professor Dr. Tipa Thep-Ackrapong)

Reader



(Mr. Michael Rudy)

Reader

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

(Associate Professor Supha Panjaroen)

March, 2005

ACKNOWLEDGEMENTS

First of all, I would like to thank the CAT Telecom Public Company Limited, for generously granting me a master's degree scholarship and a leave of absence.

I am most grateful to Asst. Prof. Dr. Tipa Thep-Ackrapong, the chair of the Master of Arts degree in English, for her guidance, insightful comments and willingness to help.

My sincere gratitude is extended to Dr. Saengchan Hemchua, my master's project adviser, whose generous assistance, helpful encouragement and insightful advice made this study possible.

I would like to thank Mr. Michael Rudy, my reader, for his valuable comments and suggestions.

My thanks also go to all my teachers, friends and relatives, who have helped me at different stages of my education.

Finally, I would like to express my gratitude to my father and my mother for the love, support and education they gave me.

Bancha Nakprakhon

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Research Questions.....	3
1.3 Scope of the Study	3
1.4 Significance of the Study.....	4
1.5 Definition of Terms.....	4
1.6 Overview of the Study.....	6
2 REVIEW OF THE RELATED LITERATURE.....	7
2.1 General Vocabulary Acquisition.....	7
2.1.1 Vocabulary Size.....	7
2.1.2 Vocabulary Acquisition	9
2.1.3 Academic Vocabulary.....	13
2.1.4 Coxhead's Academic Word List.....	15
2.2 Corpus Study.....	17
2.2.1 Corpus and Word List.....	18
2.2.2 Corpus in the Teaching of Languages and Linguistics..	19
2.3 Related Literature of Academic Words.....	25

Chapter	Page
3. METHODOLOGY.....	29
3.1 Analytical Tool.....	29
3.1.1 Analytical Framework.....	29
3.1.2 The Concordance Program.....	30
3.2 Procedure.....	31
3.2.1 Data Collection.....	31
3.2.2 Corpus Creation.....	32
3.2.3 Data Analysis.....	33
4. FINDINGS.....	35
5. CONCLUSION AND DISCUSSIONS.....	43
5.1 Conclusion.....	43
5.2 Discussions.....	44
5.3 Applications.....	48
5.4 Limitations of the Study.....	49
5.5 Suggestions for Further Studies.....	50
REFERENCES.....	52
APPENDIX An Academic Word List.....	57
VITAE.....	64

LIST OF TABLES

Table	Page
1 A sample of the Concordance program.....	31
2 Word types, word tokens and coverage of the AWL in the corpus of political reports	35
3 The 50 most frequent academic words from the AWL in the corpus of political reports	37
4 Word families in the AWL that did not occur in the corpus of political reports	40
5 Coverage of the AWL and GSL in the corpus of political reports	41

LIST OF FIGURES

Figure	Page
1 Sample concordance lines for <u>barriers</u>	21
2 Academic words and GSL words in the sample news.....	39

CHAPTER 1

INTRODUCTION

1.1 Background

How many words are there in English? Scholars try to answer this question, and they have a range from 600,000 to two million words (Schmitt, 2000 and Nation, 2001). One of the largest dictionaries of English *Webster's Third New International Dictionary* contains about 54,000 word families (Schmitt, 2000). With this huge number of vocabulary, the idea of learning all of them is impractical. According to Nation (2001), academic words cover ten per cent of a typical academic text, while the first 2,000 or high-frequency words cover 76 per cent of the text. The rest 14 per cent are technical words and low-frequency words. However, several researchers seem to agree that second and foreign language learners (SL/FL) need to concentrate first on the high frequency words of the language (Nation, 1991; Schmitt, 2000 and McChathy, 2002). For these high frequency words, West's General Service List (GSL) of 2,000 words is an efficient list of basic English words because it covers about 80 percent of a text (Carter & McCarthy, 1988; Nation, 1990; Schmitt and McCarthy, 1997 and Coxhead, 2000).

For learners of English for academic purposes (EAP), knowing only basic 2,000 words is not sufficient because spoken and written texts in the academic domain contain words at the 80,000-word level (Nation, 2000). Thus, learning words next to the basic 2,000 word-level is considered a wise strategy (Nation, 1991 and Coxhead & Nation, 2001). For this reason, Nation (1990) has attempted to find

shared vocabulary that most frequently occur in academic texts. It is believed that this group of words can assist students with the learning of academic purposes and to have enough foundation prior to the study of specific subjects. Nation, (1990) has then developed the *University Word List* (UWL), containing about 800 word families. He claims that the list covers around 8.7 per cent of academic texts. However, the list has its weakness as the criteria for the selection of words are not clear, and the size of the data is small (Coxhead, 2000).

With a computer-assisted method, Coxhead (2000) has then developed a list of 570 academic word families called *Academic Word List* (AWL) (see appendix) from a corpus of 3,500,000 words taken from academic texts and academic journals in New Zealand. She claims that the words account for 10 per cent of all academic texts found at the college level. The academic words can be used as a threshold for the teachers in class.

In fact, academic words, such as *area*, *contribute*, *establish* and *undergo* can also be found in different text types (i.e. newspaper, fiction, nonfiction.), but the coverage varies from genre to genre. Nation (2001) compared the coverage of academic words in fictions, conversations, and newspapers. He found that academic words accounted for approximately 1.7 per cent in fictions, 1.9 per cent in conversations, and 3.9 per cent in newspapers. This is much lower than the coverage of academic words in the academic corpus, which is about 10 per cent.

Since academic texts contain a considerable number of academic words, they are sometimes viewed as an uninteresting area for students when compared to other text types such as literary works or newspaper articles (Hirvela, 2001). According to Jordan (1997), the tone of the academic texts is serious, impersonal and

formal, rather than colloquial or conversational. Hence, learners may enjoy more when reading some other text types and learn academic words unconsciously. Besides, newspapers contain relatively short texts and are easily available to both learners and teachers. When learners are exposed to this kind of text, they will experience academic words necessary for academic reading and writing. Therefore, newspapers, particularly political reports, were of the researcher's interest. As such, this study aimed to explore the coverage of Coxhead's Academic Words in political reports. It was hoped that if they contained fair amount of academic vocabulary, then they would be an alternative material for academic vocabulary learning and teaching.

1.2 Research Questions

1. What percentage of the words in the AWL is in political reports?
2. What are the 50 most frequent academic words in the AWL used in the political reports?

1.3 Scope of the Study

A corpus of 122,811 words taken from political reports from www.nationmultimedia.com from January 1, 2001 to December 31, 2003 was used as the data. The analysis was in the following aspects:

1. The number of Coxhead's AWL words in the corpus of political reports.
2. The top 50 AWL words in the corpus of political reports.

1.4 Significance of the Study

This study aimed to investigate the number of academic words in political reports in a local English newspaper in Thailand. The study is beneficial as follows:

1. It will reveal the source of academic vocabulary.
2. Political reports would be an alternative teaching material for teachers and learners of English for Academic Purposes (EAP).

1.5 Definition of Terms

1. Academic words

Academic words such as *sustain*, *adjust*, *comprise*, *derive*, etc are words that are common in different kinds of academic texts. In this study, Coxhead's (2000) *Academic Word List* (AWL) which comprises 570 word families (see appendix) was used as the analytical framework. These words are the results of a study of 3,500,000 words corpus taken from textbooks, academic journals, journal articles from the World Wide Web, and laboratory manuals.

2. Word types

Word types are different word forms. A word which is repeated many times is counted only once. For example, a composition of 250 words consist of words such as *the*, *a*, *and*, *in* and *that*. Thus, the number of word types is certainly fewer than 250. In the sample sentence "If students become discouraged and quit reading, there can only be one outcome. If, however, they can be supported and encouraged, there is every possibility of success." (Coady, 2002, p.234), there are 27 words in total, but contains 22 word types because *there*, *and*, *can*, *be*, and *if* occur more than once and are counted as one word.

3. Word token

Word tokens refer to the occurrence of the total number of word forms. That is, the individual words occurring more than once in a text are counted each time they are used. For example, in the previous two sample sentences, there are 27 word tokens, because the repeated word *there*, *and*, *can*, *be*, and *if* are counted twice.

4. Political reports

Political reports refer to online news on politics, both national and international, available at www.nationmultimedia.com. The web page provided direct links to the politic archives which were kept in a text format. Most of the reports were written by local Thai people, and were concerned mostly with Thai politics.

5. Corpus

A corpus refers to a collection of texts which can be in the form of written texts or recorded speech. It can be used as a starting-point of linguistic description or as a means of verifying hypotheses about a language (Crystal 1991). In this study, a corpus was created by taking political reports from www.nationmultimedia.com. It comprised 122,811 words of political reports.

6. Concordance program (CP)

A concordance program is a computer program that helps language learners and teachers to create concordance lines of any text which has previously been electronically stored. Most CPs can perform more tasks other than creating concordance lines. They can usually produce an index, help search for words, or help us in the investigation of style, vocabulary distribution, grammatical form, rhyme

schemes etc. In this study, the CP was used to search for academic words in corpus of political reports.

1.6 Overview of the Study

This study analyzed academic vocabulary in corpus of political reports in the following manner. The literature of academic words, vocabulary acquisition, and corpus study are reviewed in Chapter 2. The chapter also discusses some applications and the importance of corpora in language study. Chapter 3 looks at the analytical framework or the *Academic Word List* (AWL), and the concordance program. Chapter 4 is devoted to the results of the study. Chapter 5, the last chapter of the research, discusses the results of the analysis in detail and shows how the AWL words found in corpus of political reports differs from Coxhead's AWL. It then concludes by discussing the pedagogical implications of the findings and the suggestions for further studies.

CHAPTER 2

REVIEW OF THE LITERATURE

There are three main areas in this chapter: general vocabulary acquisition, corpus study and related literature on academic vocabulary. The first area surveys the vocabulary size and general issues in vocabulary acquisition. It then discusses academic vocabulary, followed by Coxhead's *Academic Word List* in detail. The second area is devoted to corpus linguistics and the use of corpora in various aspects, including corpora in vocabulary teaching. Finally, it presents research on corpora and academic vocabulary, the AWL in particular.

2.1 General Vocabulary Acquisition

This section provides a survey on vocabulary size and vocabulary acquisition both in general and English as a second language learning.

2.1.1 Vocabulary size

According to Richards (2000), the size of English language in printed medium has a wide range from 400,000 to 1,000,000 million word types. *Webster's third New International Dictionary*, the largest dictionary of English, contains around 114,000 word families (Nation, 2001). McCarthy (2002) has analyzed the five million word CANCODE (Cambridge and Nottingham Corpus of Discourse in English) corpus of spoken English and a 5-million-word written sample of the Cambridge International Corpus. He has found 30,000 – 50,000 word forms in the corpus. He also estimates

that there would be considerably more in written texts, about 80,000 word forms. Nation and Waring (1997) estimate that an English native speaking graduate has a vocabulary size of about 20,000 word families. Richards (2000) suggests that EFL/ESL learners should learn these 20,000 word families in order to build native-sized vocabulary.

Hill and Laufer (2003) argue that vocabulary size in English as a second language learners, graduates of high school and university students of English as a Foreign Language, ranges between 1,500 to 4,000 word families. Yodnil's (1980) study of word recognition of first-year students at Srinakharinwirot University, Prasanmit, revealed that the students knew an average of 1,700 words. The figure is much less than an English native graduate mentioned above (18,000 words lower). Therefore, the idea of studying words according to learners' purposes is sensible (Jordan, 1997, Coxhead, 2000 and Hill & Laufer, 2003).

For this reason, Coxhead and Nation (2001) divide the vocabulary of English into four groups as follows:

1. High frequency words. High-frequency words are the 2,000 most frequent words that occur in any type of text. They cover around 80 per cent of most texts. They include function words such as pronouns, prepositions, auxiliary verbs and numbers. Some content words are also included, such as *government*, *students*, *place*, etc. The list was originally created by Michael West in 1953 and has been used as a model for a high-frequency list.

2. Technical words. The words in this group differ from one subject area to another (i.e. *inflation* in finance, *byte* in computer science, *electromagnetic field* in

physics). For any subjects, technical vocabulary consists of around 1,000 words, which cover about five percent in most texts.

3. Low frequency words (i.e. *potential, obese, maternal*). In general, this group consists of around 87,000 different word families (A word family is a headword with its inflected and derived forms. For example, *indicate, indicated, indicating, indicative, indication, indications, indicator, and indicators* are counted as one word family with the headword *indicate*). These low frequency words may occur once or twice and not appear for a long time.

4. Academic words (i.e. *abandon, primary, transfer*). This group consists of 570 word families that occur in a wide variety of academic texts. The coverage of the AWL is around 8.5 – 10 percent in academic texts; therefore, they are considered an important threshold for learners with academic purposes. The AWL is discussed in section 2.1.3.

2.1.2 vocabulary acquisition

In this part general vocabulary acquisition and vocabulary in second language acquisition are discussed.

In vocabulary acquisition, there is a distinction between productive vocabulary and receptive vocabulary. According to Nation (1990), the receptive vocabulary means learners are able to recognize a word and recall its meanings. Productive vocabulary involves the process of receptive learning plus the ability to use words appropriately both in writing and speaking. If the goal of learners is to understand texts or understand lectures, a receptive knowledge is sufficient. If learners need to use all language skills, they need the productive vocabulary of 2,000

base words and larger receptive vocabulary. Generally, productive vocabulary is smaller than receptive vocabulary. Hatch and Brown (1995) claim that the point of receptive and productive knowledge is that there are different ways to know a word. That means knowing a word, which is considered sufficient in one situation, may not be sufficient to another. However, they state that knowledge about words may include information that may cause learner to choose not to produce them. Sometimes people avoid using some words for social or other reasons. That is, knowing vocabulary involves the image the speakers want the listeners to have of him or her.

Laufer (1997) explains that receptive vocabulary is larger than productive vocabulary and the size difference between them increases with age. She states that vocabulary growth proceeds in different ways for ESL and EFL learners. Intermediate and advanced EFL learner's productive vocabulary size is closer to their receptive vocabulary than is the case with ESL learners, while ESL learners have large receptive vocabulary. It can be said that one of the main goals in vocabulary learning is to close the gap between productive and receptive vocabulary size.

In vocabulary learning and teaching, there is a concept of intentional learning and incidental learning. Klien (1986) suggests that there are two types of second and foreign language learning: tutored and untutored learning. The former involves instructions and the presentation of materials and exercises, while the latter is free from guidance such as everyday communication. The untutored learning occurs when learners stay in the second language environment. This is in accordance with Hatch and Brown's (1995) idea, as they define intentional learning or direct learning as being planned, or intended by teachers or students. Incidental learning is the type of

learning that is a byproduct of doing or learning something else. They explain that intentional learning helps in learning of words, but most of the words learned are from incidental learning, largely from reading. However, Gardner (2004) proposes that there are many factors involved in incidental vocabulary acquisition during reading such as the type of texts, the number of times learners encounter certain types of words and the amount of prior vocabulary knowledge they need to learn during reading experience.

Hill and Laufer (2003) explain that if learners pay careful attention to the word's pronunciation, grammatical, category, meaning and semantic relation to other words, they are likely to be able to use the word appropriately. McCarthy (1990) proposes the idea of mental lexicon: the process of the mind in organizing words. This process consists of three elements of mental lexicon: input, storage and retrieval. First, the input refers to linguistic input such as spoken data or a stream of sound. For the student of a second or foreign language it is spoken and written language that plays important roles. The sources of inputs may be informal, where vocabulary is acquired unconsciously. The learner may learn words from radio and television, films, pop music, international use of language such as at airports. Second, the storage is a process in which words are organized to one another. For example, when the word *honey* is presented, the learners may associate it with *money* because the mind stores words according to the spelling pattern. This also applies to sound and meaning patterns. Third, the retrieval refers to the ability to retrieve the knowledge of a word when required. In other words, the retrieval process occurs when language learners can actively use a particular word when it is needed, without too much mental searching.

According to Shostak (2004), teaching vocabulary can be categorized as two general approaches: definitional and contextual. In the definitional method, there are several steps to work with. That is, teachers have students look up words in a dictionary, write down or memorize definitions, use the words in sentences, find synonyms, and then complete worksheets or take quizzes. Students can gain a specific meaning and interesting information about the history of the word. In the contextual approach, teachers ask students to explain the meaning of a word by looking at its semantic properties, syntactic cues or by examining typographic clues from charts, graphs, or pictures. This approach can also be useful, but the context clues alone may provide only a partial meaning of a word and occasionally may even be misleading.

Ellis (1997) proposes that learners do not learn words individually, but at first they learn words in phrases. They discover that words in a phrase occur in sequences. For example, the phrase *how do you do* is learned because it occurs quite often and has its meaning. Then learners notice that they can produce phrases beginning with *how*. Ellis explains that learning a new word is to recognize it as a word and enter it into our mental lexicon. This works as a process, that is, vocabulary is learned through its syntactic properties, word structure and form, and word class. Thus, it is necessary for learners to be exposed to a great deal of the language in order to be able to use native-like language.

This section has introduced some of the basic concepts that are important in vocabulary acquisition. Sizes of vocabulary have been demonstrated, along with the concepts of receptive and productive vocabulary. Some of the vocabulary teaching and learning aspects have also been mentioned. Next, the survey of academic

vocabulary is presented, beginning with some background of academic vocabulary followed by the central theme of this study, Coxhead's *Academic Word List*.

2.1.3 Academic vocabulary

The idea of compiling a word list for academic vocabulary has been attempted by several researchers. According to Jordan (1997), there have been two types of frequency counts. One type consists of a core vocabulary of about 2,000-3,000 words (e.g. Michael West's general Service List in 1953), and counts words outside this basic list. The other type counts only those words in the university texts that cause difficulties for overseas students. Nation (1990) reported on the compilation of university word lists and how these were obtained to form University Word List of almost 740 headwords (Nation 1990). Such word list can be used as a guide and focus for teachers in different activities and as a checklist for student to acquire. Nation (2001) notes that 'technical words,' or specialist vocabulary, are usually considered the responsibility of the subject teachers: the words are closely connected with learning the subject and may present conceptual difficulties.

The next level of vocabulary below special words is variously called semi-technical, sub-technical or academic, and consists of words which occur across a number of disciplines. The nature of the academic vocabulary, as it relates to graduate research, is discussed by Martin (Nation & Coxhead, 2001). She divides the academic vocabulary into three groups, appropriate for teaching/practicing with research graduates: (1) the research process, (2) the vocabulary of analysis, and (3), the vocabulary of evaluation, as follows:

1. The research process vocabulary is primarily verbs and nouns and is presented in a context which discusses the five steps of research: formulating, investigating, analyzing, drawing conclusions and reporting results.

2. The vocabulary of analysis includes high-frequency and two-word verbs needed in order to present information in an organized sequence, e.g. *consist of, group, result from, derive, base on, be noted for*. Verbs of analysis are grouped in a number of semantic sets, e.g. *consist of, be composed of, contain* and *be made up of*.

3. The vocabulary of evaluation includes adjectives and adverbs that occur in reviews, critiques and some reports, e.g. *exhaustive, controversial, coherent, indispensable, comprehensive, distinctive, pervasive, substantive, significant* and *insignificant*.

Jordan (1997) notes that there can be different attitudes towards the teaching/learning of academic vocabulary. One attitude is that students need to be aware of and learn the general meaning of the words first. If they want to know a specialist meaning and the use of the words, the words need to be learned in its context, as this is the important aspect students will have when studying in higher education.

Corson (1997) suggests that as academic vocabulary is largely from Latin or Greek, and learning these languages would be useful for learning the vocabulary. He also argues that learners should be encouraged through the use of productive academic vocabulary, that is, through formal talks, discussions based on texts, writing summaries and reviewing the literature of the topic.

This section has introduced academic vocabulary in several aspects. We can see that academic vocabulary has been explored as early as in 1950s and can be

categorized into different sub-groups according to its functions. The next section will discuss in detail the Coxhead's *Academic Word List*, which is used as the framework in this study.

2.1.4 Coxhead's Academic Word List

The *Academic Word List* (AWL) was created by Coxhead (2000) at the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. The list contains 570 word families (see Appendix 1) which were selected according to the range and frequency of the words. Altogether, if words in each family are counted, there are up to 3,000 words in the list. The process in obtaining the AWL involved using a corpus of academic English in a wide range of academic texts from a variety of subject areas. The academic corpus contains journal articles, book chapters, course workbooks, laboratory manuals, and course notes. The texts were selected according to whether they were representatives of the academic genre in that they were written for an academic audience. With approximately 3,500,000 running words, the corpus is divided into four faculty sections: Arts, Commerce, Law and Science. Each of these faculty sections contains approximately 875,000 running words. Each faculty section is divided into seven subject areas of approximately 125,000 running words.

There are four types of words that are excluded from the *Academic Word List*. The first group are words in the first 2,000 words of English from West's General Service List (GSL) as the basic vocabulary any learner should have before starting to learn academic vocabulary. The next type is words that occur in fewer than four faculty sections of the Academic Corpus. That is, the words must occur in more than

15 of the 28 subject areas of the Academic Corpus. This criterion is set to avoid a narrow range. Also, technical or specialist words often have a narrow range and are excluded from the list. The third group is proper nouns, the names of places, people and countries. The last group is Latin forms such as *ibid*, *et al*, *etc*, and *i.e*.

Regarding the importance of the AWL, Coxhead and Nation (2001) state that academic vocabulary is common in many different academic texts, and not so common in non-academic texts. They point out that the AWL is a primary goal for learners who want to take up academic study in English. Another goal of the development of the AWL is that it could be used by teachers as part of a program preparing learners for tertiary level study or used by students working alone to learn the shared words in academic texts. In addition, AWL covers about 10 per cent of academic texts. This is a high percentage because in any text, the first 1,000 words in the GSL cover about 80 percent

Coxhead (2000) suggests several ways in teaching and learning the AWL. First, she recommended the learners be exposed to academic environment in different aspects, such as reading academic texts and listening to academic lectures and discussions. Second, the written and spoken texts should not be too difficult for learners, with no more than about 5 per cent of the running words in the texts being new words for the learners. Finally, she recommends the direct learning of the AWL such as word cards or intensive study of short academic texts.

This section has discussed the AWL, particularly on how it was created and the criteria in word selection. The focus is on its high coverage in academic texts, about 10 per cent in the corpus. The next section will explore the field of corpus study in more detail.

2.2 Corpus Study

A corpus is a collection of linguistic data, either written texts 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). A corpus can be thought of as a collection of texts gathered according to particular principles for some particular purposes such as a corpus of newspapers, a corpus of scientific novels, a Shakespeare corpus, etc. Some corpora, for example, the British National Corpus and COBUILD, which have several million sentences, and over 100 million words, are often in electronic format. A corpus contrasts with an archive because an archive is a collection where all kinds of texts are stored simply because each individual text is of interest in itself (see Crystal, 1991, Tan, 2002 and Nation, 2001 for more details).

To explore a corpus the researcher needs a software tool called a concordance program. This program enables linguists and researchers in applied linguistics to study the lexical, syntactic, semantic, and stylistic patterns of a language. Sinclair (1991) points out that the concordance program is a tool that has been widely used by linguists and literary researchers. It provides a list of the occurrences of either a particular word, or a part of a word or a combination of words in context and it is drawn from a text corpus, which is presented in context. It also provides a list of words occurring in a text arranged in alphabetical order. As opposed to the index, the concordance provides the context in which the word occurs in a given text. Thus, all examples are listed with the search item in the middle of the screen and some contexts to the left and right of it.

2.2.1 Corpus and word list.

Discussions of vocabulary learning involves an important question as to ‘How many words are the best words to learn first?’ The attempts to deal with this question can be referred back to the 1930s. Carter and McCarthy (1988) describe how Ogden, and later Richards, worked on Basic English in the early 1930s. Basic English consists of 850 words and was believed to be adequate for expressing general ideas. However, according to Carter and McCarthy, (1988) there were some problems with Basic English. Of all the 850 words, there were over 12,425 meanings attached to them. Many verbs such as *smoke* and *walk* were missing, and many everyday phrases such as *goodbye* and *thank you* were not included in the list.

In 1931, Thorndike’s published a list of the important 20,000 words. The list obtained from over 200 sources, containing about 5,00,000 words. Actually, the list rated the first 10,000 words from one to twenty, according to the frequency and range of its occurrence. However, the list had its flaws as some words, such as proper names, had lower rating than they should have, while words used in books about history and chemistry had higher rating than they should have (Fries & Traver, 1950.)

In the 1920s and 1930s, Palmer proposed a list of 3,000 words that consist of vocabulary of 1,000 words and another 2,000 words. He also worked with Michael West in the work that lead to the General Service List (GSL) in 1953. However, like Basic English, the GSL also had problems: there was no collocational information provided alongside the words, the concept of coverage was not fully developed, and the frequency alone does not necessarily make it useful for learners (Nelson, 2000).

Despite these criticisms, the influence of the GSL has continued to the present day. Carter and McCarthy (1988) term the GSL as “one of the most innovative

examples of foreign language pedagogy and lexicometric research in this century”

(p.9). The publication of texts for students and also many dictionaries has been influenced largely by this list.

However, McEnery (2003) argues that later word lists are the product of the important development in the computer which taken place in the 1950s and 1960s. Vocabulary also came into focus in the register analysis which began in the 1960s. Then the important factor for the research in lexis came with the ability of the computer to give access to a large amount of data such as present days corpora, which contain several million words, can be explored and retrieved information within a matter of seconds.

The next part will examine the roles which corpora play in the study of language. The importance of corpora to language study is the importance of empirical data that enables linguists to make objective, rather than subjective statements, which are based on the individual's perception of language.

2.2.2 Corpus in the teaching of languages and linguistics

Corpora have changed the way in which linguists can look at language (Sinclair, 1991). A linguist who has access to a corpus, or other (non-representative) collections of machine readable texts can look at all the examples of a word or phrase from many millions of words of texts in a few seconds. Dictionaries can be produced and revised quickly, thus providing up-to-date information about language. Also, definitions can be more complete and precise since a larger number of natural examples are examined.

Sinclair (1991) states that examples from corpora can be organized into more meaningful groups for analysis. Using corpus and concordance lines can show the context of the word alphabetically. Thus, it is possible to see all instances of a particular collocate together. Also, a corpus has its role in dictionary building as it enables lexicographers to look at new words entering the language, or existing words changing their meanings, or the balance of their use according to genre. The ability to show word combinations rather than individual words means that phrases and collocations are more systematically analyzed than in the past.

Concordance examples from a corpus are important in language learning as they expose students to the kinds of sentences that they will encounter when using the language in real life situations. The next figure presents the sample look of a concordance of *barriers*.

Figure 1 Sample concordance lines for barriers

1 f administrative and physical barriers after 1992/93 could, in theor
 2 India has imposed formidable barriers against D10 1010 4 the en
 3 le of passing these impinging barriers and J08 0390 2 entering t
 4 impossible. Breaking down barriers and creating an environment of
 5 20 5 Despite countless barriers and disappointments, A14 0630
 6 they tend to create further barriers and make it m less and less li
 7 ook for ways to break brewing barriers and set new industry standard
 8 ere erm er tariffs and trade barriers and what have you erm in vario
 9 itoring checks when moveable barriers are erected at each end of the
 10 frise, those sharp stakes and barriers around the K16 0990 5 for
 11 f language and other cultural barriers. As Brend Shehu, a Nigerian f
 12 itish Rail to replace manned barriers at Chitts Hill, Lexden, and Ch
 13 rriers (shelves) or vertical barriers (bulk- heads) are not present
 14 armarked to receive automatic barriers could now be fitted with safe
 15 et, Kelvedon, with automatic barriers covering only half the carria
 16 at. Moving down there, market barriers. Er and that basically erm en
 17 For example, effective mine barriers from Florida J71 1690 4 t
 18 they were immensely popular. Barriers had to be put around their pi
 19 ot so long that reproductive barriers have developed. Palmer (1984)
 20 mpact on the removal of such barriers in the absence of associated

Figure 1 shows the word *barriers* in different contexts. The graphical look of the word provides authentic examples and words occurring to the left and to the right of the target word.

Apart from being a source of empirical teaching data, corpora can be used to look critically at existing language teaching materials. Many textbooks contain only invented examples and their descriptions are based upon intuition or second-hand data (O’Keeffe & Farr, 2003). This is why corpora come into play. They emphasize the importance of corpora as “an influence in language pedagogy” (p.412). Thus, corpus exploration can provide reliable linguistic evidence for language learners. However, to work with corpora, both teachers and learners must know how to properly use instructional materials such as concordance software and dictionaries.

Corpora have also been used in the teaching of linguistics. Kirk (1994) required his students to base their projects on corpus data which they must analyze in concordance with a model such as Brown and Levinson's politeness theory or Grice's co-operative principle. In this way, Kirk is using corpora not only as a way of

teaching students about variation in English but also to introduce them to the main features of a corpus-based approach to linguistic analysis.

A further application of corpora is a role in computer-assisted language learning and teaching. Ooi's (2002) work at National University of Singapore has looked at the language of Shakespeare corpus through the World Wide Web. In this approach, she analyzed the word *love* through concordance lines to see how it behaves in the corpus. She also looked at students' writing corpora to analyze whether native speakers have any confusion in using words such as *discrete* and *discreet*. She strongly recommends using corpora in language learning because linguistic evidence reflects on the nature of language. Lin (2002) studied the use of the word *it* from two corpora: a 300,000 word corpus of essays written by natives students and a corpus of 160,000 words of essays from the writing of the students (Chinese learners) in Hong Kong Polytechnic University. She found some interesting results from the corpora as the Chinese learners clearly misuse the word *it* in various contexts. She concluded that such comparison study would not only help learners identify and understand their mistakes, but also motivate them to learn more independently.

McCathy (2002) examined advanced level vocabulary by using two corpora as the data source: the five-million word CANCODE spoken corpus and CIC (Cambridge International Corpus). He found that the vocabulary in frequency of 6,000 – 10,000 band have different meanings. A word such as *spine*, which means part of the human body, occurs in different contexts. The meanings of *spine* can be part of a book where the binding is attached or main vertical item in a network. Another example is that the 6,000-10,000 band has *clasp*, *cling*, *grip*, *grasp*, but no

grab, which is put into a lower band, occurring more than 300 times. The reason is that *grab* is a general word or has the same meaning as those words. In addition, the word is assumed to be known by the learners, otherwise, it will be put into the advanced level as a new word. McCarthy called this circumstance “the mismatch between frequency and psychologically associated networks”(p.22).

Schmitt (2002) illustrated how corpus evidence can be used by teachers to teach and assess vocabulary. In teaching vocabulary, he presents two ways in which corpora can be used: deductive and inductive teaching approaches. In a deductive teaching approach, corpus evidence can be used to better inform teachers about language structures they are presenting, and to provide clearer authentic examples of language elements. The examples of using concordance lines between the words *think of* and *think about* are given. In this case, the teacher can provide the nature of the two words, which one is preferred over the other, or what kind of words usually occur after them. Likewise, in an inductive approach, corpora can be used to provide linguistics data from which learners can learn rules and language structures by themselves, and learn language structures or words they are interested in.

St.John (2001) examined whether a parallel corpus and a concordance would be an appropriate tool to a teaching program of German at the beginners' level in an unsupervised environment. In this instance, a beginner student of German was asked to find satisfactory answers to unknown vocabulary and formulate appropriate grammar rules for himself using the parallel corpus and concordancer as the only tools. He found that these tools could be of great benefit for beginners. Also, using corpora and a concordancer can be motivating and rewarding not only for the learner

but also for the teacher. For the teacher, these tools can provide clear examples and lead to answers to some lexical questions students have in mind.

In this section we can see how a number of areas of language study have benefited from using corpus data. McEnery (2003) summarized the main important advantages of corpora as follows:

1. **Access:** Once a set of texts has been gathered and placed in a corpus it can be made available to researchers all over the world. Moreover, it provides non-native speakers with the same possibilities of study as native speakers.
2. **Broad scope of analysis:** Corpus analysis allows researchers to statistically test their hypothesis and carry out research that can produce precise and reliable results.
3. **Pedagogic:** Authentic data from the corpora could be a motivation to learners in a classroom. Several researchers have provided positive feedback over the use of corpora (Ooi, 2002; Aston, 2002 and Schmitt & Zimmerman, 2002).
4. **Reliability:** The reason is that computers are more reliable in the process of data analysis. Additionally, the empirical evidence presented by a corpus of authentic texts can present strong evidence and the frequency of particular items.

In conclusion, this section has shown that corpus study in modern days depends largely on computer technology. Corpora of several million words can provide reliable and authentic data for learning and teaching English language. The last section will provide reviews of works related to academic words.

2.3 Related Literature of Academic Words

There have been several studies related to academic words. As the *Academic Word List* is relatively new, most of the studies were conducted after the year 2000, as follows:

Schmitt and Zimmerman (2002) examined the ability of 106 graduate and undergraduate nonnative-English-speaking students from the United States and United Kingdom to produce derivative forms of words from the AWL. The population of the study was expected to produce the four major word classes (i.e. noun, verb, adjective, and adverb) from 16 words, such as *access*, *assumption*, *precision*, and *persistency*. They found that knowledge of one word in a family did not necessarily imply productive knowledge of other forms in that family. Schmitt and Zimmerman suggested that reading academic texts provide exposure to academic words, and would aid the acquisition of academic word derivatives. Donley and Reppen (2001) used a concordance program in a classroom to identify academic vocabulary in an academic corpus of cultural anthropology. They searched the words from the corpus that also occur in the *Academic Word List* as the target of the pretest. The results of the pretest were used to determine which academic vocabulary items would be presented in the class. In order to make the amount of vocabulary manageable, they chose 10 academic words for a three-week period. They concluded that the role of instructor be as important as the concordance program in identifying important academic vocabulary. However, the number of selected words for instruction must not be too many for learners to study at a time.

Kaewphanngam, Broughton and Soranasataporn (2002) gathered and analyzed a corpus of psychology texts. Then they tested students' comprehension of

technical and sub-technical vocabulary in the corpus. The subjects were undergraduate students majoring in psychology at the Department of Psychology and Guidance, Faculty of Education, Silpakorn University. They found that the 10 most frequent words from the AWL were: *research, participants, psychology, analysis, process, strategies, theory, task, item* and *negative*. Regarding the students' test results, they concluded that the students' performance on technical vocabulary and academic vocabulary in the psychology corpus was not at the level to understand the content of texts. Finally, They suggested that the frequency lists of vocabulary and concordance lines could be beneficial to material writers and teachers in language learning.

Poonpon, Honsa Jr. and Cowan (2001) studied the number of academic words of the AWL in the teaching materials used in English for science courses at Mahidol and Khon Kaen Universities by using the concordance software WordSmith Tools. The texts analyzed in the study were materials taken from English for science courses for first and second year science students at Mahidol University and Khon Kaen University. They found that some AWL words were found in both the first and second year Mahidol corpora. The first year students met 369 word families and in second year they encountered 498 word families. Of the latter, 349 had been encountered in the first year, and 149 were new. At Khon Kaen University, the first year students encountered 327 AWL word families, and 275 of these were met again together with 146 new words in the second year. However, 34 word families in the AWL such as *currency, parameter, subordinate, terminate, and unify* never occurred in the corpus.

Cobb (2001) designed a 13-week experimental course and on-line resources before testing the amount of high frequency vocabulary and academic vocabulary knowledge gained by the 33 students at Concordia University, Montreal. The 33 students represented a variety of first language backgrounds: two thirds of the group were speakers of Asian languages (Chinese and Vietnamese) and about one third had Romance language backgrounds (Quebec French, Spanish or Portuguese). He found that the learners could recognize the meanings of 80 per cent of the tested words at the 2,000, 3,000 and Academic levels. However, the study also revealed that individuals varied considerably. For example, some of the French speaking students scored high on the Academic list (which contains many words of Latin origin) but low on the 2,000 list (which contains many words of Germanic origin). Many of the Asian students (mostly Chinese speakers) scored high on the test of the 2,000 list but low on other lists. The results support the argument made by Corson (1997) about high Latin origin words in academic words and learning the Latin words would be of great benefit to learning academic words.

Coxhead (2000) examined the coverage of the AWL in fiction texts from a corpus of 3,763,733 words. The texts in the corpus were written more than 50 years ago and are available to the public. She found that the AWL accounted for approximately 1.4 per cent of the tokens in the fiction corpus. This is much lower than the ten per cent coverage of the academic corpus. She explained that the word families in the AWL were associated highly with academic writing. Also, the age of the fiction texts might be another reason for the low percentage because some words, such as *infrastructure*, did not appear in the corpus. In total, there were 30 word families of the AWL that were not in the fiction texts.

This section has reviewed studies related to the AWL. Many studies focused on finding the number of words in the AWL, particularly in textbooks. The results of those studies indicated that the number of academic words clearly varied according to the type of genre. It appears that one particular genre does not contain all the 570 word families, even in academic books. This means learners may need to read extensively to meet all the words

In conclusion, this chapter has shown some concepts on vocabulary size, vocabulary acquisition and the ideas of productive and receptive vocabulary. It has reviewed the early word lists and how they were obtained. The importance of computer in later word lists was also addressed, as well as the detailed framework of this study: the *Academic Word List* (AWL). Finally, studies related to academic words were reviewed. Regarding the studies of the AWL in Thailand, the focus was on the number of words in academic books at the university level, while studies conducted abroad focused on the acquisition of academic words.

The next chapter will consider the research questions noted in the first chapter: those concerning the number of words from the AWL in the political corpus and the 50 most frequent academic words from the corpus.

CHAPTER 3

METHODOLOGY

This chapter consists of two main sections: 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 description of the method used in selecting the data and a corpus creation. The procedure employed to analyze the data is also included in this section.

3.1 Analytical Tool

3.1.1 Analytical framework

Coxhead's (2000) *570-Academic Word List* (AWL) was used as the framework in this study. Created from the size of 3.5 million running words, the list has to date been the most complete and reliable (Coxhead & Nation, 2001, p.252). The 570 word families are the shared vocabulary of written academic English used in a wide variety of disciplines (28 subjects in total). Selection was based on the range and frequency, using an academic corpus of journal articles, book chapters, course workbooks, laboratory manuals, and course notes. These 570 words consist of head-words plus their inflected and derived forms, (e.g. *concept*, *conception*, *concepts*, *conceptual*) and therefore there are 3,110 word-forms altogether.

3.1.2 The Concordance program

The *Concordance* (Watt, 2004) is a program for Windows based computer. It can make wordlists, concordances, from electronic texts. Users can make wordlists - lists of all the different words that occur in a text. The wordlists can be arranged (sorted) in many different ways. Each word can be optionally accompanied by a count of the number of times it occurs, and by a note of the percentage of all words it represents.

The program can make complete concordances to all the words in the text. It can handle large texts with no limits on the size of the corpus. In addition, the *Concordance* has a graphical look and ways in which a user can view and interact with a concordance. It seems that the important feature of the *Concordance* is the ability to build a concordance on any large corpus. Table 1 provides a sample look of the *Concordance* program.

Table 1 A sample of the Concordance program.

Number	Word Frequency	Per cent
101 long	775	1.09
102 map	474	1.00
103 top	486	0.74
104 invite	453	0.74
105 perfect	425	0.65
106 associate	418	0.64
107 promise	404	0.47
108 affair	402	0.47
109 foreign	397	0.45
110 arrive	376	0.44
111 success	340	0.38
112 withdraw	339	0.35
113 bypass	322	0.33
114 intern	302	0.32
115 aside	268	0.31
116 tower	255	0.31
117 doctor	178	0.30
118. ...		
119. ...		
120...		

Table 1 provides a sample of word ranks and word frequency generated by the *Concordance*. The last column is the percentage of each word in a corpus. To have a better look, a user can sort words according to its ranks or its frequency.

3.2 Procedure

This research was undertaken in accordance with the following steps:

3.2.1 Data collection

The data used in this study were the political report corpus of 122,811 words taken from on-line database at www.nationmultimedia.com from January 1, 2002 to

December 31, 2003. There were 200 reports in total. Section 3.2.2 describes the process in creating the corpus.

3.2.2 Corpus creation

In this study, the target was to create a corpus of 100,000 words from political reports at www.nationmultimedia.com. The steps in creating the corpus are as follows:

a) Political reports from the website were copied. The homepage of the website shows links to different archives which are categorized into smaller sections, such as entertainment, sports, education, health, politics, and editorial. Then the political report texts were copied one at a time.

b) Dates and months in the reports were removed from the texts. Only headlines and the body of the news were copied into a master file. For example, there is 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* is one of the words in the academic wordlist.

c) The texts were inserted into a political report file. The corpus was fully developed.

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

3.2.3 Data analysis

In order to answer the two main research questions in this study, the researcher produced a concordance list of the corpus by using the *Concordance* program.

Research question 1: What percentage of the words in the AWL is in political reports?

In order to answer this question, the overall percentage of the AWL had to be obtained through the following steps.

- a) Finding out the total tokens of the corpus by using the option *Make Full Concordance*.
- b) The program presented total word tokens of the corpus.
- c) The number of all word tokens the *Concordance* generated from the corpus was stored.
- d) The total AWL words were calculated by using the *Concordance's Pick List* tool. This tool appended the total words of the AWL, headwords and their families. Then the AWL word tokens were obtained.
- e) The percentage of the AWL was obtained by using the following formula:

$$\text{the percentage of the AWL} = \frac{\text{total AWL word tokens (d)} \times 100}{\text{total word tokens in the corpus (b)}}$$

Research question 2: What are the 50 most frequently used academic words (AWL) in the political reports?

In order to answer this question, the frequency of each word needed to be manually compared, which can be described as follows:

- a) The frequency of each word in the AWL obtained from the procedures in the first research question was manually compared.
- b) The top 50 academic words were identified and stored.
- c) The top 50 academic words would then be reported in a table similar to table 1.

In conclusion, the need for corpus-based research in academic words has been stressed. For a large corpus consisting of over 100,00 words, it is necessary to be analyzed using computer program as a tool, as this chapter has presented under the corpus creation and the *Concordance* program sections. The next chapter will present the results of this study

CHAPTER 4

FINDINGS

This chapter presents the findings of the study according to the following research questions

1. What percentage of the words in the AWL is in political reports?
2. What are the 50 most frequent academic words (AWL) used in the political reports?

The results of the study are presented and discussed. It will first present the broader picture of the corpus of political reports such as the total word tokens and total word types. Then it will present the tables related to the questions of the study.

In order to explore the number of words in the AWL in the corpus of political reports, the word tokens of both the AWL and the entire corpus of political reports was compiled.

Table 2 Word types, word tokens and the coverage of the AWL in the corpus of political reports.

Type of words	Number of words
The AWL word families	513
The AWL word types	1,219
The AWL word tokens	7,020
Total word types in the corpus	11,703
Total word tokens in the corpus	122,811
The AWL coverage (%)	5.7

Table 2 shows word types, word tokens and the coverage of the words in the AWL in the entire corpus. The AWL word types in the corpus accounted for 10.5 % of all word types in the corpus (1,219 out of 11,703), while the AWL word tokens accounted for 5.7 % of all word tokens (7,020 out of 122,811). That is to say, the 5.7 % figure was the actual coverage of the AWL words in the corpus political reports. Although this figure is less than Coxhead's 10% coverage of academic texts, the corpus contained most of the AWL word families (413 word families or 90 % of the total 570 word families). Finally, the number of word types in the corpus (11,703 words) indicates the vocabulary size in the corpus of political reports.

In order to answer the second research question, the frequency of headwords and its families are compiled and counted as one word as presented in Table 3.

Table 3 The 50 most frequent academic words from the corpus of political reports

Word families		Frequency	Per cent
1.	constitution	403	5.74
2.	prime	347	4.94
3.	policy	169	2.40
4.	economy	133	1.89
5.	ministry	121	1.72
6.	commission	118	1.68
7.	issue	99	1.41
8.	major	89	1.26
9.	administrate	88	1.25
10.	military	75	1.06
11.	found	74	1.05
12.	academy	61	0.86
13.	media	60	0.85
14.	finance	56	0.79
15.	respond	56	0.79
16.	role	56	0.79
17.	involve	55	0.78
18.	fund	52	0.74
19.	comment	47	0.66
20.	secure	46	0.65
21.	participate	46	0.65
22.	job	46	0.65
23.	community	44	0.62
24.	seek	44	0.62
25.	conflict	43	0.61
26.	culture	43	0.61
27.	investigate	42	0.60
28.	promote	42	0.60
29.	require	42	0.60
30.	select	40	0.56
31.	debate	40	0.56
32.	strategy	38	0.54
33.	team	38	0.54
34.	authority	37	0.52
35.	process	37	0.52
36.	create	36	0.50
37.	project	36	0.50
38.	research	35	0.49
39.	corporate	34	0.48
40.	dominate	33	0.47
41.	institute	33	0.47
42.	focus	32	0.46
43.	affect	31	0.45
44.	benefit	31	0.45
45.	file	31	0.45
46.	decade	29	0.41
47.	deny	29	0.41
48.	despite	29	0.41
49.	link	29	0.41
50.	violate	28	0.39

Table 3 displays the 50 most frequent academic words in the corpus. The frequency of individual words in the families was aggregated and then presented in headwords. Note that some words may have the same frequency. It was found that the most 50 words were topic-related. Many words clearly reflected the field of politics such as *constitution*, *policies*, *ministry*, *economy*, *administrate*, *military* and *issue*. The word *constitution* had the highest frequency of occurrence (403 or 5.74 %) and clearly reflected the type of the corpus of political reports, while the word *violate* had the lowest frequency of occurrence (27 or 0.38 %).

It can be explained that “constitution” can occur in various forms (*constitute*, *constituencies*, *constituency*, *constituent*, *constituents*, *constituted*, *constitutes*, *constituting*, *constitution*, *constitutions*, *constitutional*, *constitutionally*, *constitutive*, *unconstitutional*) and cover more meanings in the political reports than *prime*. Note that the word *prime* always occurs with *minister* to constitute *prime minister* which means the leader of the government.

The proportion of the type of vocabulary in the political reports can be seen in the Figure 2.

Figure 2 Academic words and GSL words in sample news

Rakkiat seen conspiring to hide wealth

*Lead prosecutor Klanarong Chantik yesterday **summed up** a graft case against former public health minister Rakkiat Sukthana as a conspiracy to conceal his family's ill-gotten wealth.*

*"Before entering politics, Rakkiat was a lawyer and his wife Surakanya an accounting auditor. Together they worked as a husband-and-wife **team** to cover up assets, suspected to be ill-gotten earnings, when **filing** their asset statements," he said.*

*In his rebuttal, Rakkiat said the National Counter Corruption **Commission** (NCCC) was **biased** against him, which led to unfair accusations.*

*Klanarong and Rakkiat both made their **final** pitches before the **Constitution Court**.*

*The prosecution contended that Rakkiat and his wife had used their combined **legal** and accounting knowledge to conceal assets held under the names of their relatives.*

*"The asset statements had appeared to be in order, but the NCCC later uncovered many assets that had not been declared. Confronted with this **evidence**, Rakkiat admitted to concealing his assets," Klanarong said.*

In his defence, Rakkiat said many of the undeclared assets belonged to Surakanya and he had not been kept informed of them. In her testimony for the defence, Surakanya said she had to conceal some of the family's assets because she did not want her husband to lose them gambling.

*But Klanarong contested the pair had made a number of gambling trips together and had no reason to withhold information from one another. The **tracing** of the family's **financial** records also **indicated** that the two were **aware** of and had **access** to accounts they claimed to have concealed from one another, he said.*

Published on June 3, 2003

The above sample shows words in the GSL (in italics), which accounted for 191 words (74%), while the words in the *Academic Word list* (in bold) accounted for 14 words (5.5%). Other words accounted for 52 words (20.5%). The total word

types were 257 words. This proportion is valid in most academic texts (Coxhead, 2000 and Nation, 2001).

However, not all the AWL words occurred in the corpus of political reports. The AWL words which did not occur in the corpus are presented in Table 4.

Table 4 Word families in the AWL that did not occur in the corpus of political reports.

1. abstract	21. domain	41. presume
2. aggregate	22. duration	42. precise
3. analogy	23. empirical	43. publication
4. append	24. exhibit	44. qualitative
5. arbitrary	25. extract	45. random
6. assemble	26. finite	46. ratio
7. attain	27. fluctuate	47. refine
8. bulk	28. incentive	48. revise
9. cease	29. inherent	49. route
10. chemical	30. inhibit	50. sequence
11. coherent	31. innovate	51. simulate
12. commence	32. intermediate	52. undertake
13. compatible	33. interval	53. uniform
14. complement	34. intrinsic	54. unify
15. conceive	35. journal	55. unique
16. conform	36. layer	56. utilize
17. contemporary	37. modify	57. whereby
18. depress	38. paradigm	
19. deviate	39. parameter	
20. dispose	40. portion	

Table 4 shows that, of the 570 word families in the AWL, there were 57 word families (10%) that did not appear in the political corpus. However, the readers would meet most of the academic words in political reports because the rest 90% of the word families in the AWL occurred in the corpus. The disappearance of the 57 word families suggests that some words may be found frequently in a particular field, for example, in chemistry-related topics (*chemical, extract, portion*), or research

papers (*empirical, random, qualitative*). However, it is interesting to note that most of the non-apparent words are rated 8, 9, or 10 by Coxhead (2000), which means that they occur infrequently in academic texts.

The next table provides the overall picture of the corpus of political reports. The *Concordance* program compiled words in the AWL and GSL as presented in Table 5.

Table 5 Coverage of the AWL and GSL in the corpus of political reports.

Word groups		Word Tokens	Word Types	Coverage (%)
Academic Word List		7,020	1,219	5.7
General Service List				
First	1,000 words	84,423	2,430	68.7
Second	1,000 words	6,764	1,249	5.5
Others		17,584	6,805	20.1
Total		112,811	11,703	100

Table 5 shows that GSL words accounted for most of the tokens in the entire corpus. Of the total 122,811 tokens in the corpus, the GSL accounted for 98,207 tokens or 74.2 %, while words in the others group accounted for 21.1 % of all the total words, about one out of every five words in the corpus. Academic words accounted for 5.7 % coverage of the corpus. There were 90 per cent of the AWL word families (513 out of the 570) found in the corpus.

The results yielded slightly different proportion from that of Coxhead' (2000) study, that is, the large proportion was the GSL (76 %), followed by other word groups (14 %) and the AWL (10 %) respectively. Note that the 17,584 word items, which did not fall in the AWL or the GSL, may indicate that there is a large proportion of low frequency words or content specific words that the corpus of political reports contains.

It can be concluded from the results in Table 2 to 5 that the number of the academic words in the corpus of political reports was lower than those in Coxhead's (2000) study. However, this is not surprising because the academic words, as the name suggests, are naturally high in academic texts. The list obtained from the corpus of political reports also provided interesting aspects in vocabulary use in a political genre. The next chapter will discuss the results, some suggestions for applications and further studies.

CHAPTER 5

CONCLUSION AND DISCUSSIONS

This chapter answers the two research questions about the number of academic words from the AWL in corpus of political reports and the first fifty words from the AWL used in the corpus. Then the discussion, applications, limitation of the study and suggestions for further study are presented.

5.1 Conclusion

Research question 1. What percentage of the words in the AWL is in political reports?

The AWL words accounted for 5.7 % in the corpus of political reports. There were 7,020 word tokens out of the total 122,811 word tokens in the corpus. The corpus contained 513 AWL word families, accounting for 90 % of the total 570 word families. The AWL coverage in the corpus was about four percent less than that of Coxhead's (2000) study of Academic corpus (10%), and three percent less than her specific findings on the coverage of vocabulary in science discipline (9.1%).

Research question 2. What are the 50 most frequently used academic words (AWL) in the political reports?

Many words in the 50 most frequently used academic words (AWL) were related to the field of politics, for examples, *constitution, prime, policy, economy, ministry, commission* and *issue*. Of all the academic words from the AWL,

constitution, occurred the most (403 frequency or 5.74%), followed by *prime* (347 frequency or 4.94 %) and *policy* (169 frequency or 2.40%). The least frequent word in the top 50 academic word list was *violate*, which had the frequency of 27 or 0.39%.

5.2 Discussions

As noted in the introduction, the results of this study would be useful to teachers and learners of English in that they may consider political reports in particular, as alternative materials both in and off class. The findings provided important characteristics of the AWL in the political reports and should be taken into account in deciding whether this kind of texts should be used as teaching materials. It can be seen from the findings that academic vocabulary and its rank were different from Coxhead's (2000) study. She claimed that the 570 AWL word families accounted for around 10% of the total tokens in an academic corpus. After acquiring GSL words, knowing academic words should be the goal for learners wishing to study at a university where they are expected to read their subject textbooks in English. In this study, the words from the AWL had 5.7% coverage of the political reports from *The Nation* online news database. Although this figure is four percent less than Coxhead's 10% coverage of academic texts, it might be acceptable if we look at newspapers as one of effective sources of authentic texts for foreign language study. Besides, newspapers contain short texts which require relatively short time to read compared to academic articles or other types of academic papers. Thus, alternative learning materials like this can be useful in learning and teaching academic words because they can provide learners with opportunities to meet words in new contexts. In addition, online newspapers can generate great interest among

learners since it is a source of up-to-the-minute English and they are accessible to everyone.

The AWL words found in the corpus also shared two important features. First, as the topic of the text obviously has a great effect on the frequency of the words, most of them contain meanings related to a political sense. This indicates that learners would be exposed to some words from the AWL that have specific meanings in a political sense. For example, *prime* can be a noun, adjective and verb. However, in this corpus *prime* always occurred with *minister* to constitute the meaning of head of a government. The explanations for this are that the corpus was created from political context in Thailand, and the prime minister is always central to the public interest because of his duties and responsibilities as the head of the government. Therefore, it is inevitable that the word *prime minister* has been, and will always be, covered by every newspaper, including *The Nation*. It also indicates that word lists alone might not provide full understanding of the texts because individual words can associate with other words to form a totally different meaning. Thus, it seems necessary to learn academic words by comparing their meanings with those in other genres in order to productively use these words.

Second, most of the high frequency AWL words in this corpus can play an important role in politics-related vocabulary (i.e. *issue, minister, administrate and policy*). Normally, academic words tend to be overshadowed by content-specific vocabulary (Donley and Reppen, 2001, p.10). As a result, it reduces incidental learning of academic words. In this case, the balance of learning politics-related vocabulary and academic vocabulary is needed when reading some specific genres such as political reports. Without sufficient content-specific vocabulary, learners will

not fully understand the texts they are reading. Therefore, studying the list obtained from the political news will help learners understand the content of the genre and also provide them good opportunities to learn academic words.

There are also some observations on newspaper texts to be discussed here. First, reading comprehension does not always involve vocabulary knowledge alone. Background knowledge of a specific genre such as political reports is needed. This would help readers predict the content of the text, guess unknown words in context, and recognize text structure, which ultimately help them understand the text thoroughly. Thus, the reader may need some time to be familiar with this type of genre, as some political reports may not be understood without prior knowledge. Second, in reading political reports, the vocabulary represents real events in the present time and contains many rare words, while reading other genres such as novels and short stories, in contrast, may involve scene and setting not found in newspapers. Thus, some vocabulary, particularly those used in a daily basis found in reading fictions and novels, may not be found in newspapers. This is in accordance with a study by Grabe and Stoller (2002), who suggested that reading newspapers is different from reading novels and short stories in terms of types of vocabulary found in the two genres. This suggests that political reports should be applied to upper intermediate or advanced students in order that the students do not miss some important words used in everyday situations.

This study also found that words from the *General Service List* occurred most often in the corpus of political reports (80%), which is consistent with studies by Sutarsyah, Nation, and Kennedy (1994) and Coxhead (2000) which showed that GSL words accounted for 80% or more in most text types. Nation (1990, 2001) and

Coady (2002) suggest that learners of English language should master the 2,000 basic words first before moving on to other types of vocabulary. Thus, the idea of learning the GSL words should be the priority for second language learners.

The next observation is on the disappearance of the 57 word families, such as *empirical, extract, and qualitative*. Several studies have also shown that one particular genre in a limited number of words does not contain all the 570 word families (Poonpon, Honsa Jr. & Cowan, 2001 and Kaewphanngam, Broughton & Soranasataporn, 2002). This indicates that learners should read extensively in order to meet all the 570 word families. As mentioned earlier, however, long exposure to one type of text may generate boredom to the students. In this case, teacher may consider a direct learning strategy proposed by Coxhead (2000). This strategy involves using flash cards, which contain academic words and short descriptions, to introduce academic vocabulary to the students.

To sum up, several issues concerning the AWL in the corpus of political reports have been discussed. It has shown that reading political reports is worthwhile. However, the level of the learners should also be taken into account. Political reports may be appropriately applied to upper intermediate or advanced class. Finally, learners should read extensively or use direct learning strategy in order to meet all the 570 words families.

5.3 Applications

1. According to the *Concordance* program, it appears that the concordance lines are very useful for language teaching and learning. Teachers may use concordance lines in class to highlight some words in the AWL in context. Since the context is readily located by the use of a concordance program, teachers and course designers can present the learners with appropriate contexts for the use of vocabulary. In addition, the highlighted vocabulary always comes with collocates that learners can understand in different situations.

2. Learners from beginner to advanced levels can use the interactive concordance to study vocabulary on their own in a self-access center. In this case, the role of teachers is to explain how to use a concordance program before each activity.

3. The data of this study might be used for study units in an EAP reading course. Teachers can use frequent AWL words in political reports as a starting point because they are words that the students may be most familiar with. Teachers may develop indexes or glossaries and inform the students the meanings of particular words that are presented in the read texts before the reading activities.

4. Apart from political reports, the reading texts can comprise different genres such as entertainment, sports and business. The reason is that exposure to one particular genre may generate boredom and lead to familiarity with words occurring frequently only in that genre.

5. Teachers should also give attention to the 2,000 word families of the GSL because the GSL is of great value in providing coverage of the vocabulary in both specialized texts and academic texts.

6. Teachers may shift from reading to viewing news from established news agencies (i.e. Central News Network (CNN) or British Broadcasting Corporation (BBC)). The positive sides to that are, for example, students can practice listening skill and enjoy the moving pictures on the screen at the same time. However, instead of showing live reports, teachers should present recorded news which can be run over again and again to highlight particular words in the context.

In conclusion, this section provides the value of concordance lines in learning words in the AWL. Also, in addition to political reports, learners can shift to other sources such as viewing English news. Finally, teachers and learners should pay much attention to the basic 2,000 words as these words cover much of the texts regardless of its genres.

5.4 Limitations of the Study

1. In this study, the selected political reports might be regarded as one local-oriented news from only one source of English newspaper in Thailand, *The Nation* online news. Furthermore, most of the news in the corpus was written by local Thai people. Therefore, the corpus might not be as representative of political news as those collected from others such as the AP (Associated Press), CNN or BBC.

2. The results of this research will not be exactly the same as other results from similar research in the future in terms of word frequency and rank because the findings were unique in their characteristics such as political incidents in the year 2001-2003 (i.e. member of parliament election, political scandal.)

3. The findings cannot be generalized because of the small size of the corpus. A study on a bigger corpus may yield different results.

5.5 Suggestions for Further Studies

1. This study was conducted with only one genre: the political reports.

Therefore, there should be similar research with a bigger size of the political reports.

The results should be compared with wider range sections in the newspapers such as business, entertainment, editorials and information technology.

2. In this study there were a large number of low frequency words; those occur outside the range of the 2,000 word families of the GSL and the AWL. Research on this particular area would provide a clearer picture of the words learners should also pay attention to in reading newspaper texts.

3. There should be a study of collocations and the use of the words in the AWL in newspaper texts in order that learners can use the words productively.

4. There should be a comparative study of the AWL in a spoken corpus in the political reports. If the AWL words occur frequently enough in the news, it would also be an alternative source for learners apart from newspapers.

REFERENCES

REFERENCES

- Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: much more than bean counting. *TESOL Quarterly*, 35(2), 331-340.
- Candlin, C.N. (1988). *Vocabulary and language teaching*. New York: Longman.
- Carter R., & McCarthy, M. (1988). *Vocabulary and language teaching*. London: 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.
- Cobb, T. & Horst, M. (2001). Growing academic vocabulary with a collaborative on-line database. Retrieved March 22, 2004 from 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.
- (2004). An academic words list. Retrieved January 10, 2004, from <http://www.vuw.ac.nz/lals/staff/averil-coxhead/awl/awlinfo.html>.
- Coxhead A., & Nation, P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew, & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes* (pp. 252-267). Cambridge: Cambridge University Press.
- Crystal, D. (1991). *A Dictionary of linguistics and phonetics*, Oxford: Blackwell.

- Cummins, J. (2000). Academic language learning transformative pedagogy, and information technology: Towards a critical balance. *TESOL Quarterly*, 34(3), 537-615.
- Donley, M. K., & Reppen, R. (2001). Using corpus tools to highlight academic vocabulary in SCLT. *TESOL Quarterly*, 10(2/3), 7-12.
- Ellis, N.C. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt , & M. McCarthy (eds.), *Vocabulary description, acquisition and pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.
- Fries, C. C. & Traver, A. A. (1950). *English word list: a study of their adaptability for instruction*. Michigan: The George Wahr Publishing.
- 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.98-122). Cambridge: Cambridge University Press.
- Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics*, 25(1), 1-37.
- Hatch, E., & Brown, C. (1995). *Vocabulary, semantics and language education*. Cambridge: Cambridge University Press.
- Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching*, 41(2), 87-106.
- Hirvela, A. (2001). Incorporating reading into EAP writing courses. In J. Flowerdew, & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes* (pp. 252-267). Cambridge: Cambridge University Press. .
- Hyland, K. (2002). Options of identify in academic writing. *ELT Journal*, 56(4), 351-358.

- Jordan, R.R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Kaewphanngam, C., Broughton, M.M. & Soranasataporn, S. (2002). Corpus-based analysis: Guidelines for getting practical language input in materials development. Retrieved March 24, 2004 from [http://www.sc.mahidol.ac.th/sc/g/SLLT/sllt2002/ArtileSLLT\(Maurice\).doc](http://www.sc.mahidol.ac.th/sc/g/SLLT/sllt2002/ArtileSLLT(Maurice).doc)
- Klien W. (1986). *Second language acquisition*. New York: Cambridge University Press.
- Lin, L.H.F. (2002). Overuse, underuse, and misuse: Using concordancing to analyse 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.
- Laufer, B. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt , & M. McCarthy (Eds.), *Vocabulary description, acquisition and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press
- Mauranen, A. (2003). The corpus of English as lingua franca in academic settings. *TESOL Quarterly*, 3, 513-525.
- McCarthy, M. (1992) *Vocabulary*. Oxford: Oxford University Press.
- McEnery, T. (2003). Corpus linguistics. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 448-463). Oxford: Oxford University Press.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

- Nation I.S.P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary description, acquisition and pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.
- Nelson, M. (2000) A corpus-based study of business English and business English teaching materials. Unpublished PhD Thesis. Manchester: University of Manchester. Retrieved February 19, 2004 from <http://www.kielikanava.com/thesis.html>
- O’Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, 37(3), 389-418.
- Ooi, B.Y.B. (2002). From Shakespeare to Hungarian EFL writing: Using WWW corpora to motivate student learning. In M. Tan (Ed.), *Corpus studies in language education* (pp. 163-177). Bangkok: IELE Press.
- Poonpon, K., Honsa, S.Jr., & Cowan, R. A. (2001). The teaching of academic vocabulary to science students at Thai universities. Retrieved May 10, 2004, from <http://www.sc.mahidol.ac.th/sc/g/SLLT/2001>.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- 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.
- Schmitt, N., & Zimmerman, B. C. (2002). Derivative word forms: What do learners know?. *TESOL Quarterly*, 36(2), 145-169
- Shostak, J. (2004) The value of direct and systematic vocabulary instruction. Retrieved May 10, 2004, from http://www.sadlier-oxford.com/pdf/9147-9_VW_WhitePaper_Vol7.pdf.

- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford : Oxford University Press.
- St.John, E. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning & Technology*, 5 (3).185-203.
- Stubbs, M. (1996). Corpus evidence for norms of lexical collocation. In G. Cook, & B. Seidhofer (Eds.), *Principle & practice in applied linguistics* (pp. 245-270). Oxford: Oxford University Press.
- Sujan, P. (2000). *An analysis of vocabulary used in Rajabhat Institute Mahasarakham English majors' writing*. Unpublished thesis, Srinakharinwirot University, Bangkok, Thailand.
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal*, 25(2), 34-50.
- Tan, M. (2002). *Corpus studies in language education*. Bangkok: IELE Press.
- Thurstun, J., & Candlin, N. C. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, 17(3), 267-280.
- Yodnil, S. (1980). A Study of vocabulary recognition of the first year students at Srinakharinwirot University, Prasarnmit. Unpublished thesis, Srinakharinwirot University, Bangkok, Thailand.
- Watt, R. J. C. (2004). Concordance programme. Retrieved January 12, 2004, from <http://www.rjcw.freeseve.co.uk/>

APPENDIX

An Academic Word List (AWL)

The AWL of 570 headwords was developed by Coxhead (2000). It was used as a framework to investigate the number of academic words in political reports in this study.

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

considerable	consist	constant
constitute	constrain	construct
consult	consume	contact
contemporary	context	contract
contradict	contrary	contrast
contribute	controversy	convene
converse	convert	convince
cooperate	coordinate	core
corporate	correspond	couple
create	credit	criteria
crucial	culture	currency
cycle	data	debate
decade	decline	deduce
define	definite	demonstrate
denote	deny	depress
derive	design	despite
detect	deviate	device
devote	differentiate	dimension
diminish	discrete	discriminate
displace	display	dispose
distinct	distort	distribute
diverse	document	domain
domestic	dominate	draft
drama	duration	dynamic
economy	edit	element
eliminate	emerge	emphasis
empirical	enable	encounter
energy	enforce	enhance
enormous	ensure	entity
environment	equate	equip
equivalent	erode	error
establish	estate	estimate
ethic	ethnic	evaluate
eventual	evident	evolve
exceed	exclude	exhibit
expand	expert	explicit
exploit	export	expose
external	extract	facilitate
factor	feature	federal
fee	file	final
finance	finite	flexible
fluctuate	focus	format
formula	forthcoming	foundation
found	framework	function
fund	fundamental	furthermore
gender	generate	generation

globe	goal	grade
grant	guarantee	guideline
hence	hierarchy	highlight
hypothesis	identical	identify
ideology	ignorance	illustrate
image	immigrate	impact
implement	implicate	implicit
imply	impose	incentive
incidence	incline	income
incorporate	index	indicate
individual	induce	inevitable
infer	infrastructure	inherent
inhibit	initial	initiate
injure	innovate	input
insert	insight	inspect
instance	institute	instruct
integral	integrate	integrity
intelligence	intense	interact
intermediate	internal	interpret
interval	intervene	intrinsic
invest	investigate	invoke
involve	isolate	issue
item	job	journal
justify	label	labour
layer	lecture	legal
legislate	levy	liberal
licence	likewise	link
locate	logic	maintain
major	manipulate	manual
margin	mature	maximise
mechanism	media	mediate
medical	medium	mental
method	migrate	military
minimal	minimise	minimum
ministry	minor	mode
modify	monitor	motive
mutual	negate	network
neutral	nevertheless	nonetheless
norm	normal	notion
notwithstanding	nuclear	objective
obtain	obvious	occupy
occur	odd	offset
ongoing	option	orient
outcome	output	overall
overlap	overseas	panel
paradigm	paragraph	parallel

parameter	participate	partner
passive	perceive	percent
period	persist	perspective
phase	phenomenon	philosophy
physical	plus	policy
portion	pose	positive
potential	practitioner	precede
precise	predict	predominant
preliminary	presume	previous
primary	prime	principal
principle	prior	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	reveal	revenue
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	supplement
survey	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
volume	voluntary	welfare
whereas	whereby	widespread

VITAE

VITAE

Name: Mr. Bancha Nakprakhon
 Date of Birth: March 6, 1974
 Place of Birth: Amphur Bankruad, Buriram
 Address: 11 Moo 3 Tambon Bankruad, Amphur Bankruad,
 Buriram 31180
 Present Position: Telecommunication officer 5
 CAT Telecom Public Company Limited, Amphur
 Muang Nontaburi 11000

 Educational Background:
 2005 Master of Arts (English)
 Srinakharinwirot University, Bangkok
 1995 Bachelor of Arts (English)
 Buriram Rajabhat Institute, Buriram
 1992 Certificate of Science (Mathematics and Science)
 Bankruad Wittayakarn School, Buriram