CORPUS-BASED ANALYSIS OF ENGLISH VOCABULARY IN WEATHER FORECAST TEXTS AND A DESIGN OF A SAMPLE DICTIONARY

SIWARAK PAENGPHO

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ARTS IN APPLIED LINGUISTICS
(ENGLISH FOR SCIENCE AND TECHNOLOGY)
FACULTY OF INDUSTRAIL AND EDUCATION
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2011

KMITL-2011-ED-M-252-190

หัวข้อวิทยานิพนธ์	การวิเคราะห์คำศัพท์จากคลังข้อมูลภาษาอังกฤษในเอกสารพยากรณ์อากาศและการสร้างพจนานุกรมตัวอย่าง
นักศึกษา	นางสาวศิวารักษ์ แพงโพธิ์
รหัสประจำตัว	51064205
ปริญญา	ศิลปศาสตรมหาบัณฑิต
สาขาวิชา	ภาษาศาสตร์ประยุกต์-ภาษาอังกฤษเพื่อวิทยาศาสตร์และเทคโนโลยี
พ.ศ.	2554
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ. ดร. จิราภา วิทยาภิรักษ์

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คำศัพท์จากคลังข้อมูลภาษาในเอกสารพยากรณ์อากาศ จำนวน 555,818 คำ (Tokens) ซึ่งแบ่งออกเป็น 2 ประเภท คือ 1) พยากรณ์อากาศ และ 2) เอกสารอุตุนิยมวิทยา ช่วงปี พ.ศ. 2546 ถึง 2552 เครื่องมือที่ใช้ในการวิเคราะห์คำศัพท์ได้แก่ โปรแกรมจำแนกศัพท์ตามความถี่ (WordSmith Tools) รุ่นที่ 3.0 โดย Scott (1998) ค่าสถิติที่ใช้ในการศึกษาครั้งนี้คือ ค่าความถี่และค่าร้อยละ คำทั้งหมดที่พบจากการศึกษาครั้งนี้แบ่งออกเป็น 4 กลุ่ม ได้แก่ 1) กลุ่มคำศัพท์พื้นฐานซึ่งจำแนกด้วยบัญชีคำศัพท์พื้นฐาน (GSL) ซึ่งจัดทำโดย Bauman and Culligan (1995) 2) กลุ่มคำศัพท์วิชาการ ซึ่งจำแนกด้วยบัญชีคำศัพท์วิชาการ (AWL)โดยการรวบรวมของ Coxhead (2000) 3) กลุ่มคำศัพท์เทคนิคซึ่งระบุจากปทานุกรมอุตุนิยมวิทยาและ 4) กลุ่มคำอื่นๆที่ไม่สามารถจัดอยู่ใน 3 กลุ่มข้างต้น

ผลจากการวิเคราะห์คำศัพท์พบว่า คำศัพท์ที่ใช้ในเอกสารพยากรณ์อากาศนั้นในกลุ่มคำศัพท์พื้นฐานมีความถี่ค่อนข้างสูง จากคำที่ศึกษาทั้งหมด 13,172 คำ ประกอบด้วยคำศัพท์พื้นฐานร้อยละ 76.37 คำศัพท์วิชาการร้อยละ 5.86 คำศัพท์เทคนิคร้อยละ 5.30 นอกจากนี้ยังพบคำศัพท์คำอื่นๆที่ไม่ได้ปรากฏอยู่ในคำศัพท์สามกลุ่มข้างต้นอีกร้อยละ 12.58 การศึกษาด้านระบบคำพบว่ามีทั้งระบบคำเปิด และ คำปิด ซึ่งมีคำบางคำเป็นได้ทั้งคำศัพท์พื้นฐานและคำศัพท์เทคนิคด้านอุตุนิยมวิทยา ในส่วนของคำปรากฏร่วมนั้นพบว่ามีทั้งคำปรากฏร่วมประเภทไวยากรณ์ และคำปรากฏร่วมประเภทเนื้อหา นอกจากนี้ยังพบว่ามีคำศัพท์เทคนิคด้านอุตุนิยมวิทยาหลายคำที่เกิดจากการรวมกันของคำศัพท์พื้นฐาน คำศัพท์วิชาการ หรือคำในกลุ่มอื่น ในกรณีของคำย่อที่พบมีทั้งหมด

6 แบบ คือ 1) คำย่อที่มาจากการตัดคำท้าย 2) คำย่อที่ใช้เฉพาะอักษรแรกของชื่อเต็ม 3) คำย่อแบบการย่นคำ 4) คำย่อที่มาจากการผสมของอักษรต้นของชื่อเต็ม 5) คำย่อที่มาจากภาษาต่างประเทศ และ 6) สัญลักษณ์

ผลที่ได้จากการศึกษาความถี่ของคำศัพท์เทคนิคและ คอนคอร์แดนซ์ (Concordance) ได้นำมาสร้างเป็นตัวอย่างคำที่บรรจุอยู่ในพจนานุกรมตัวอย่าง ซึ่งประกอบด้วย คำหลักภาษาอังกฤษ ความหมายภาษาไทย การออกเสียงคำไทย ไวยากรณ์ คำเหมือนภาษาไทยและภาษาอังกฤษ ตัวอย่างการใช้คำภาษาอังกฤษ และภาพประกอบ

Thesis Title	Corpus-based Analysis of English Vocabulary in Weather Forecast Texts and a Design of a Sample Dictionary
Student	Miss Siwarak Paengpho
Student ID.	51064205
Degree	Master of Arts
Program	Applied Linguistics-English for Science and Technology
Year	2011
Thesis Advisor	Assoc. Prof. Dr. Jirapa Vitayapirak

Abstract

This study aimed to analyze vocabulary used in English weather forecast texts. The samples of this study were the English weather forecast texts which were divided into two main types: 1) weather forecasts and 2) meteorological documents during the year 2003-2009 in order to compile a Weather Forecast Corpus (WFC) which contains 555,818 words (tokens). The research instrument of the study was the concordance software "Wordsmith Tools Version 3.0" (Scott, 1998). The statistical results in this corpus study were frequencies and percentages. All words in this study were divided into four groups. 1) General vocabularies were classified with the General Service List (GSL) created by Bauman and Culligan (1995). 2) Academic vocabularies were classified with the Academic Word List (AWL) gathered by Coxhead (2000). 3) Technical vocabularies were identified by meteorological glossaries and 4) Miscellaneous.

The results of the study revealed that the group of general vocabulary was the highest frequency group in this study. According to the corpus finding, the total 13,172 word types in the Weather Forecast Corpus consist of the general vocabularies (76.37% of tokens), the academic vocabularies (5.86% of tokens), the technical vocabularies (5.30% of tokens), and the other vocabularies (12.58 % of tokens). The open classes and closed classes were both found in the corpus. Some general words were found to be used as technical vocabularies. Many multi-word terms were created from the combination of general, academic and others vocabularies. There

were six types of abbreviation: 1) clippings, 2) initialisms, 3) acronyms, 4) contractions, 5) substitutions, and 6) symbols.

These corpus findings in terms of word frequencies and concordance were used to develop sample entries for the proposed bilingual (English and Thai) meteorological dictionary. It offers English headword, Thai definition, Thai and English synonym, Thai pronunciation, grammatical information, English example of usage and illustration.

ACKNOWLEDGEMENT

First of all, I wish to express my sincere gratitude to all my lecturers in M.A. in Applied Linguistics at King Mongkut's Institute of Technology Ladkrabang (KMITL) for the great knowledge and valuable experience.

I would like to express my sincere gratitude and deep appreciation to my advisor, Assoc. Prof. Dr. Jirapa Vitayapirak who gave me valuable guidance, precious time, beneficial comments, patience in editing my paper, encouraged and supported me through this study. Also, I would like to express my gratitude to my thesis committee who devoted their time and advice.

I wish to extend my sincere appreciation for the help of two my former instructors at Thai Meteorological Institute - Mr. Sanong Sarapan and Mr. Tanya Thongnunui for giving me valuable recommendations. I would also like to thank all official staffs in Thai Meteorological Department who participated in user needs analysis. I am indebted to my bosses, Mr. Mongkol Sritakool and Mr. Thanat Janfaeng who allowed me to study.

I give my sincere appreciation to Mr. Mana Tosatjawong, my senior MA. student who helped me with computer software and my beloved friends, Mr. Anannet Teanhom and Ms. Chindrarat Taengkonk who have helped me in many ways at various times and places.

My greatest personal thanks is to Mr. Somsak Theppichitsamut, my closed and beloved friend who has been patient with my sensitive temper and always been at hand whenever I need help in finishing the thesis.

Lastly, I wish to express my heartfelt thanks to my beloved parents, Mr. Satid and Mrs. Samorn Paengpho, including my eldest sister, Capt. Saowanit Paengpho and younger brother, Mr. Teerawat Paengpho for their unfailing emotional support, sympathy, admiration, confidence, and faith over the years in helping me focus on my study.

Siwarak Paengpho

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES AND GRAPHS

Figure **Page**

Graph

LIST OF ABBREVIATIONS

AMS	American Meteorological Society
AWL	Academic Word List
BNC	British Nation Corpus
COBUILD	Collins Birmingham University International Language Database
EAP	English for Academic Purposes
EBE	English for Business and Economics
EOP	English for Occupational Purposes
ESP	English for Specific Purposes
ESS	English for Social Sciences
EST	English for Science and Technology
GSL	General Service List
IPA	International Phonetic Alphabet
KWiCFinder	Keyword in Context Finder
LGP	Language for General Purposes
LOB	Lancaster-Oslo-Bergen
LRF	Long Range Forecast
LSP	Language for Specialized Purposes
MD1	Weather News
MD2	Weather Report
MD3	Meteorological Journal
MD4	Meteorological Magazine
MD5	Meteorological Textbook and Manual
MD6	Other Meteorological Documents
MRF	Medium Range Forecast
MWEs	Multiword Expressions
OCR	Optical Character Recognition
OV	Other Vocabulary
POS	Part Of Speech
SRF	Short Range Forecast

LIST OF ABBREVIATIONS (CONTINUED)

TESP	Teaching English for Specific Purposes
TMD	Thai Meteorological Department
TMI	Thai Meteorological Institute
TNPs	Technical Noun Phrases
TV	Technical Vocabulary
UN	The United Nations
WMO	World Meteorological Organization
WF1	Weather Advisory
WF2	Daily Forecast
WF3	Shipping Forecast
WF4	Weekly Forecast
WF5	Monthly Forecast
WF6	Time of Sunrise-Sunset and Moonrise-Moonset Forecast
WF7	Water Level Forecast
WFC	Weather Forecast Corpus
WR1	Daily Weather Summary
WR2	Weekly Weather Summary
WR3	Monthly Weather Summary
WR4	Annual Report
WST	WordSmith Tools

CHAPTER 1

INTRODUCTION

1.1 Statement and Significance of the Problem

Nowadays, English plays a crucial role in supporting modern organizations in Thailand. English is a very important language mainly because it bridges people of the world together. Thai Meteorological Department (TMD) is a member of The World Meteorological Organization (WMO) which is a specific organization of the United Nations (UN) for weather, climate and water. The vision of TMD is "Aspiring to excellence in meteorology at the international level" (TMD, Online: 2006). English for weather forecast is, therefore, an important communication tool that enables TMD to become one of the international meteorological service providers.

It is generally acknowledged that English for weather forecast is a kind of English for Specific Purposes (ESP). One important problem for ESP learners seems to be English lexical knowledge in reading, writing, and translating. Vocabularies are the core of learning procedure in terms of their meanings and usages. When ESP learners encounter difficult vocabularies (technical terminology, for instance), they often search for them in dictionaries. In the past, it was very difficult to compile a technical dictionary because many specialized terms had to be manually selected and carefully defined by experts in that particular field which took a lot of time and effort. In addition, the technical terms were not selected from authentic or real-life texts.

As technology changes, computers have been used in vocabulary studies and in lexicography for several decades (Landau, 2001). A corpus-based study is widespread in learning specialized vocabulary and making technical dictionary from authentic texts. In other words, the corpus-based study is another way for using computer readable forms of texts for the purpose of linguistic research and making dictionaries. There are several advantages in using specialized corpora including searching, sorting, retrieving, and calculating linguistic data with high speed and accuracy (Mallikamas, 1991).

Modern Dictionaries have been revolutionized by the introduction of corpus-based techniques and they are now usually based upon huge corpora of English, from which words, forms, spellings, meanings and grammatical behavior are extracted, thus allowing lexicographers to appeal directly to the observed facts of language (Trask, 1999: 166). However, little previous

research had been carried on the design of specialized corpora. None of the specialized corpora so far has been applied for the bilingual (English and Thai) meteorological terminologies. In Thailand the bilingual meteorological dictionaries have not been well-developed and very few references in this area are available. No meteorological dictionary for Thais was based on corpus research and compiled by the corpus-based analysis which met the needs of the users.

This research is an attempt to apply the corpus-based study for compiling a specialized corpus for weather forecast which enables us to select English weather forecast texts in order to design a sample of meteorological dictionary suitable for Thai users.

1.2 Goals and Objectives

The purpose of this study is to compile a corpus of English for weather forecast, and analyze English vocabulary from the texts to design a sample of technical dictionary with the following four main objectives:

1.2.1 To establish the Weather Forecast Corpus (WFC).

1.2.2 To analyze word classes, collocations, compound nouns, and abbreviations.

1.2.3 To classify word lists in the corpus into general, academic, and technical vocabularies.

1.2.4 To design a sample of meteorological dictionary based on the technical vocabularies found in the Weather Forecast Corpus.

1.3 Scope and Limitations of the Study

There are six limitations in this study.

1.3.1 The vocabulary inputs of the study were English weather forecasts texts collected from three main sources, i.e. 1) newspapers: Bangkok Post; 2) websites: Thai Meteorological Department (TMD)'s website, World Meteorological Organization (WMO)'s website, and other international websites; and 3) documents in TMD working units.

1.3.2 The samples were collected and grouped into two main groups: 1) Thailand weather forecasts and 2) meteorological documents used in TMD working units from the year 2003 to 2009. The group of weather forecast consists of two main categories, i.e. types of media (printed, electronic, and both electronic and printed) and period of time (short range forecast,

medium range forecast, and long range forecast). Thailand weather forecasts can be divided into seven sub-categories: Warning Advisory (WF1), Daily Forecast (WF2), Shipping Forecast (WF3), Weekly Forecast (WF4), Monthly Forecast (WF5), Sunrise-Sunset and Moonrise-Moonset Forecast (WF6), and Water Level Forecast (WF7). The group of meteorological documents consists of six categories: Weather News (MD1), Weather Report (MD2), Meteorological Journal (MD3), Meteorological Magazine (MD4), Meteorological Textbook and Manual (MD5) and Other Meteorological Documents (MD6).

1.3.3 The General Service List (Bauman and Culligan, 1995) and Academic Word List (Coxhead, 2000) were used as the guidelines for identifying the general and academic vocabulary respectively.

1.3.4 The guidelines for identifying the technical vocabulary were taken from 'the Online Glossary of Meteorological Terms' (English and Thai) written by Thai Meteorological Department (2007), 'the Glossary of Meteorological Terms written by Thai Meteorological Department (1979), and 'the Online Glossary of Meteorology' written by American Meteorological Society (AMS) (2000).

1.3.5 The guidelines on the abbreviation analysis were based on 'Abbreviations Dictionary' written by Sola, Stahl, and Kerchelich (1995).

1.3.6 Only Thai weather forecasts were used in this study. Since the Weather Forecast Corpus (WFC) was collected from Thai technical writers, it could cause some possible problems of English usage.

1.4 Process of the Study

There are five main parts in the process of this study: literature review, data collection, data analysis, results and discussions, and conclusions and suggestions.

1.4.1 Literature Review

The related literature of research involving vocabulary, corpus-based methodology and lexicography were studied.

1.4.2 Data Collection

1) The weather forecast texts were collected, i.e. weather forecasts (55.6 %) and meteorological documents (44.4 %).

2) In the corpus compilation, all printed weather forecasts were scanned and electronic texts were downloaded and saved as plain text files (*.txt).

1.4.3 Data Analysis

1) A computer software program, WordSmith Tool (WST) Version 3.0 (Scott, 1998) was used to examine tokens, word types and calculate type/token ratio and word frequency of the whole corpus. All texts were categorized into the types of vocabulary. After that they were presented in terms of percentage of word occurrence and word frequency.

2) From the word lists, the identification of general and academic vocabulary are based on the General Service List (GSL) provided by Bauman and Culligan (1995) and Academic Word List (AWL) provided by Coxhead (2000) respectively.

3) From the word lists, the identification of technical vocabulary is based on 'the Online Glossary of Meteorological Terms' (English-Thai) written by Thai Meteorological Department (2007), 'the Glossary of Meteorological Terms written by Thai Meteorological Department (1979), and 'the Online Glossary of Meteorology' written by American Meteorological Society (2000).

4) The collocations were classified into lexical and grammatical collocations by putting emphasis on technical compounds.

5) Based on the Abbreviations Dictionary (Sola, Stahl, and Kerchelich, 1995), the abbreviations were examined.

1.4.4 Results and Discussions

The results were presented as follows:

1) The overall statistical results were presented in terms of number of tokens and word types, type/token ratio, and word frequency list.

2) The word classes were classified into verbs, nouns, adjectives, and adverbs.

3) Collocations were analyzed in terms of lexical and grammatical collocations by putting emphasis on technical compounds.

4) Abbreviations were listed.

5) The frequency and percentage of occurrence of the general, academic, and technical vocabulary were presented.

6) The sample of meteorological dictionary was designed.

7) Discussion was provided.

1.4.5 Conclusions and Suggestions

Conclusions and suggestions for further studies were given.

1.5 Definitions of Terms Used

Academic word list (AWL) the list of academic words that are selected for tertiary level students who need to read and write academic prose. The AWL contains 570 word families based on a 3,500,000 tokens corpus of academic English.

Collocations the way in which words are frequently used with another words in weather forecast texts

Concordance the collection of the occurrences of a word-form, each in its own textual environment (Sinclair, 1991).

Corpus the collection of language in natural text both of written and spoken that have been gathered in electronic form.

General Service List (GSL) the list of the basic vocabulary of English in order of frequency, and the words chosen to be the basic vocabularies for learners of English as a second language. It was first created by West in 1953. The version used in this study was adapted by Baumann and Culligan in 1995. It contains almost 2,300 words.

Lemmatization the process of finding the base forms of a lexical unit, which is known as the *lemma.* For example, a word like *used* in a written text, we first have to find the lemma *use* to look up information about the word in a dictionary. The word-forms *use*, *uses, used, using*, and probably *to use*, will conventionally be lemmatized into the lemma *use.*

TMD	the abbreviation of Thai Meteorological Department, a member of the World Meteorological Organization which carry on meteorological administrations and managements.
Token	the individual occurrence of a word or all words form in a text, for example, 'go', 'went', 'gone', 'going' consists of four tokens or running words.
Weather Forecast	English report on likely weather details for a period of time in the future in Thailand.
Word Frequency Lists	the list of vocabulary items in a corpus. This information can be shown in either alphabetical order or frequency order.
WMO	the abbreviation of The World Meteorological Organization that is a specific organization of the United Nations (UN) for meteorology (weather and climate), operational hydrology and related geographysical sciences.

1.6 Anticipated Outcomes

This study is aimed to use findings of the English vocabulary used in weather forecast texts with which a corpus-based study was used for analysis. Hopefully, this study can help to develop meteorological dictionary at the Thai Meteorological Department (TMD). It can assist English teachers, meteorological students, meteorologists, meteorological officers, lexicographers, linguists, and translators, and people who are involved or are interested in weather forecast or meteorology in the description of meteorological language. The findings of this research will provide language source in English for Specific Purposes (ESP) with particular reference to writing texts and teaching English at Thai Meteorological Institute (TMI).

CHAPTER 2

LITERATURE REVIEW

This chapter is divided into four main parts. The first part begins with weather forecast concepts which consist of definition, importance, and types of weather forecasts. The second part reviews related research on weather forecasts. The third part describes word classes, collocations, abbreviations and vocabulary. In the last part, corpus lexicography and technical dictionaries are explored.

2.1 Weather Forecasts Concepts

2.1.1 The Definitions of Weather Forecasts

The weather forecast is a summary of the meteorological parameters (i.e. cloud, temperature, pressure, wind, rain, visibility, humidity) that have been obtained at a specified place and time. It is the specific data reported on whether it is a climatological, rainfall, or synoptic station (Dunlop, 2001; Forthofer and Pedro, 2007). According to the definition in American Meteorological Society's Online Glossary (AMS, 2000), weather forecast is defined as an assessment of the future state of the atmosphere with respect to precipitation, clouds, winds, and temperature. Such assessments are usually made by government or private meteorologists, often using numerical simulations. Such simulations are the result of representing the atmosphere mathematically as a fluid in motion. Saslow and Mongillo (1985) describe that weather forecast is the weather information taken by the application of science and technology to predict the state of atmosphere for an advance time and in a given location. In short, it is a way to predict the weather and sometimes the word 'weather forecast' is called by forecasters as 'weather report' or 'weathercast' (Sievert, 2005). The weather forecasts tell people what kind of weather to expect in the days ahead and the states of the natural disasters are such as flood, storms, storm surge, thunderstorm, and so on.

2.1.2 Importance of Weather Forecasts

Human lives often depend on the weather conditions. Nowadays, the weather forecast is often used to be a topic of daily conversation because the weather affects every aspect of human live. People around the world need to know what the weather will be like today or tomorrow in

order to know the possibilities of weather conditions. In short, the weather forecasts are very importance for people around the world. There are two main reasons as follows:

Firstly, the weather forecasts strongly influence in daily living plan of people. People need to know what the weather will be like today or tomorrow in order to prepare for engaging in other travel. Better information about the weather allows people to make better decision such as preparing themselves for outdoor activities or, simply planning what to wear and whether to take an umbrella, outside sporting, taking a raincoat, communicating, washing and drying clothes in the air. Moreover, the forecast of natural disasters (e.g., hurricane, storm, and flood) can save lives of people. These are advantages of the weather forecasts.

Secondly, the weather forecasts affect work such as air-traffic, energy, marine, agriculture, construction, transportation, trade, and business of travelling. For instance, the significance of weather forecast with air-traffic will concern with aviation. A pilot needs to know about the present weather or next hour weather to prevent an accident in the sky from strong wind or poor visibility. Moreover, weather forecasts aid oil company staff in deciding when to perform tasks (Sripada et al., 2002) in order to prevent oil products from bad weather. Sometimes the prevention in the disruption of system of communication and transportation need to rely on the weather forecast as well. Therefore, the weather forecasts have important roles for human around the world in this era.

2.1.3 Types of Weather Forecasts

The weather forecasts are normally grouped according to criteria of Thai Meteorological Department (TMD, Online: 2007). They can be divided into three main groups: short range forecast, medium range forecast, and long range forecast.

2.1.3.1 Short Range Forecast (SRF)

Generally short range forecast is a weather forecast for up to a few hours (usually not more than 72 hours) into the future. It can be divided into three sub-division groups as follows:

2.1.3.1.1 Nowcast

Nowcast is used to describe the weather forecast which is usually not more than 3 hours. For example, a weather forecast for up to a few hours is called a nowcast such as Warning News.

2.1.3.1.2 Very Short Range

Very short range is used to state the weather forecast from 3 hours to 12 hours such as Warning News.

2.1.3.1.3 Short Range

Short-range is used to refer to the weather forecast from 12 hours to 72 hours into the future such as Warning News, Daily Forecast, and Shipping Forecast.

2.1.3.2 Medium Range Forecast (MRF)

A medium range forecast is a weather forecast for predictions from 72 hours to 10 days in advance time such as 7-days Forecast.

2.1.3.3 Long Range Forecast (LRF)

A long range forecast is a weather forecast that extends from 10 days up into the future such as Agricultural Forecast, Monthly Forecast, and Season Forecast.

2.2 Related Research

Bowker and Pearson (2002) studied specialized terms, collocations, grammar, and style of weather forecast. They compiled a small corpus on weather forecasts from *Scientific American* data and used corpus-based techniques in order to look up frequencies of word, headwords that are specialized terms, and their collocations.

The results were that the written style of weather forecasts was very short containing a very few sentences. Weather forecast text has a different writing style from the normal Language for General Purposes (LGP) style as shown in Figure 2.1.

1. a thunderstorm.	High near 68F.	**Winds** N to NE 10 to 15 mph.
2. f a thunderstorm.	Low near 56F.	**Winds** SW 5 to NE 10 mph. Chanc
3. a thunderstorm.	High near 69F.	**Winds** N to N 15 to 20 mph. Chanc
4. sun and clouds.	High near 78F.	**Winds** S 10 to 15 mph. Frida
5. thunderstorms.	Low around 62F.	**Winds** E to SE 5 to 10 mph.
6. d thunderstorms.	High near 76F.	**Winds** SE 5 to 10 mph. Chanc
7. ; Mostly cloudy.	High near 66F.	**Winds** W 10 to 15 mph.
8. Partly cloudy.	High around 68F.	**Winds** W to SW 10 to 15 mph.
9. ; Partly cloudy.	Low near 53F.	**Winds** N to NW 5 to 10 mph.
10. Partly cloudy.	Low around 57F.	**Winds** N to NE 10 to 15 mph.
11. Partly cloudy.	Low around 58F.	**Winds** S to SW 10 to 15 mph.
12. Partly cloudy.	Low around 50F.	**Winds** W to SW 5 to 10 mph.

Figure 2.1 Concordance for 'Wind/Winds' (Bowker and Pearson, 2002: 37-38)

This concordance is generated for the term 'wind/winds' based on these short extracts. A number of observations about the style of weather forecasts can be made. The style of weather forecasts is very sparse. Sentences are very short, and there are no unnecessary words such as articles (e.g. 'the' or 'a'), nouns (e.g. 'temperature') or even verbs (e.g. 'blowing'). In this corpus, the term 'winds' is used only in the plural form, they are not found in the singular. Each individual forecast is very formulaic: it begins with a description of cloud cover and precipitation, moves on to temperature and ends with a description of the winds. The order never changes. Even within each of these descriptions, the information is presented in a systematic way: winds are always described first in terms of direction and then in terms of speed. Bowker and Pearson (2002: 37-38) points out that articles such as 'the' or 'a', nouns such as 'temperature', and verbs such as 'blowing' are not necessary.

Reiter and Sripada (2003) studied the use of time phrases in weather forecasts. Their study was based on a corpus analysis technique where align phrases from the forecast text with data were extracted from a numerical weather simulation. The results were the parallel text-data corpora which are valuable for investigating lexical semantic and pragmatic issues, and can help shed valuable light on the fundamental question of how words relate to the non-linguistic words; especially the results are many time phrases in weather forecasts. Here are some examples:

Table 2.1 Time Phrases in the Weather Forecasts (Reiter and Sripada, 2003)

Time	Phrases Used in the Weather Forecast
00.00 am.	by late evening/around midnight/ by midnight
03.00 am.	tonight/ in early hours/after midnight
06.00 am.	overnight/in early morning/by early morning
09.00 am.	by midday/during morning/ by (mid) morning
12.00 am.	by midday/around midday/ by midday
15.00 pm.	by mid afternoon/ in mid afternoon/ by mid afternoon
18.00 pm.	by evening/ in early evening/ by early evening
21.00 pm.	by evening/during night/ by (mid)

Srisirirungrot (1998) also studied the form of weather forecast news presentation and its strategies in presenting news in disastrous weather including its role. She found that weather

forecast news generally had particular aspects especially, when disaster occurs. In addition, bad weather will be added in the weather forecast news. She concluded that the major role of the weather forecast news presented by mass media is to inform about the weather. Moreover, the weather forecast news takes more roles in the warning and instructing.

2.3 Word Classes

Word classes are well known as part of speech. They are crucial for any grammatical explanation. In English, linguists agree on the need to recognize the following word classes: noun, verb, adjective, preposition, adverb, determiner and conjunction. Curzan and Adams (2006) divide word classes into two major classes: open class categories and closed class categories.

2.3.1 Open Class Categories

Open class categories or also known as lexical word classes are open-ended (Curzan and Adams, 2006). There are four major open class categories: nouns, adjectives, verbs, and adjectives.

2.3.1.1 Nouns

Nouns are words representing people, animals, places, things, or abstractions such as 'water', 'mountain', 'dust', 'city', 'child', or 'children'. There are many kinds of nouns, i.e. collective noun, count noun, uncounted noun, mass noun, proper noun, singular noun, plural noun, and compound noun.

2.3.1.2 Adjectives

Adjectives are words which are used to describe the state of a person or thing, or give special detail about them. Usually, it is used to modify a noun or pronoun and can tell about kind or number (such as *high, low, odd, strange, heavy*). Here are examples:

The *strange* noise included a *high* note.

It made an *odd* sound.

2.3.1.3 Verbs

Verbs are action words. It is used with a subject to say what someone or something does or what happens to them. Lexical verbs are such as *'call'*, *'sit'*, and '*work*'. For example:

The producer *called* her director.

2.3.1.4 Adverbs

Adverbs are words that add detail about the action, event, or situation mentioned in a clause. They are used as a modifier of a verb, an adjective, another adverb, a phrase, and a clause. It refers to the manner, place, or time. They are as follows:

Adverbs of manner	e.g. *actively, aggressively, fast, carefully*
Adverbs of place	e.g. *upstairs outside outdoors everywhere*
Adverbs of time	e.g. *today, yesterday, later, now, last year, never, often*

2.3.2 Closed Class Categories

Closed class categories or grammatical word classes are defined as closed systems which result from the elements of grammatical classes (Curzan and Adams, 2006). There are six major categories: prepositions, conjunctions, pronouns, complementizers, determiners, and auxiliary verbs.

2.3.2.1 Prepositions

Prepositions are words that are used to link nouns, pronouns and phrases to other words in a sentence and helps in indicating location, direction, time, duration, manner, and other relationship e.g. *at, in, on,* and *to* as the following examples:

She went to bed *at* midnight.

We will have our lunch *on* the train.

2.3.2.2 Conjunctions

Conjunctions are words that link words, phrases, and clauses as in the following examples:

We eat with fork *and* a spoon.

I can ride a bicycle *but* I can't ride a horse.

Conjunction can be classified into three types:

1) Coordinating conjunctions are a connector for the word or phrases of two nouns or two prepositions (such as *for, and, but, nor, or,* and *yet*).

2) Subordinating conjunctions connect two main clauses i.e. a main clause and a dependent or subordinate clause (such as *because, before, after, when, although, unless, if, while, in order that, as long as,* and many others).

3) Correlative conjunctions focus on paired conjunctions (such as *either/or, neither/nor*, and *not only/but also*).

2.3.2.3 Pronouns

Pronouns stand in for nouns or noun phrases; can be subjects, objects, and complements. They are classified as closed class. Pronouns are nouns representing unique entities e.g. *Paris, Earth* or *John*, as separated from common nouns that focus on a class of entities e.g. *city, planet* or *person*. They can be subject, object, and complements. Proper nouns are classified as closed classes. There are actually five categories of pronouns as follows:

Personal pronouns like *he, she, they*, or *we*.

Indefinite pronouns like *anyone, someone, everyone, no one, something, neither, either, another, both, all, most, some, whoever, whomever*, or *whatever*.

Interrogative pronouns like *who/whom/whose* and *what/which*.

Demonstrative pronouns like *this, that, these,* and *those*.

Relative pronouns like *who/whom/whose, that, which, whoever/whomever*, or *whichever*.

2.3.2.4 Complementizers

In English, two most common complementizers are *that* and *which*, that function as complementizers when used to advise a clause that serves as the object of a verb and some verbs such as *'say'* and *'think'*, they can require a full clause to perfect their meaning in a sentence as in the following examples.

His close friend *said that* the weather is good now, but I *think that* it's suppose to rain this afternoon.

2.3.2.5 Determiners

Determiners encompass the class of function words which suggest noun phrases. They are used to indicate determinacy (such as *a, an, the*) quantity (such as *some, many, all*), number (such as *one, two, first, second*), pragmatic functions (such as *this, that, these, those*). Greenbaum (1996) divides determiners into three types, there are as follows:

- Predetermines like *all, both, twice, half, double,* and *ten times*.
- Central determiners like *these, this, my, your, which, whose, which ever, some, any,* and so on.
- Postdeterminers like *six, third, three, seventeenth, another, last, next, many, several, few*, and *much*.

2.3.2.6 Auxiliary Verbs

Auxiliaries are verbs which show grammatical rather than lexical information, and are used along with lexical verbs denoting events. In English, auxiliary verbs can be divided into two main classes: primary and modal verbs (Greenbaum, 1996: 153-156).

- Primary verbs like *be, do, have*
- Modal verbs like *can, could, may, might, must, shall, should, will, and would.* They express necessity, obligation, or possibility.

2.4 Collocations

Collocation is the way in which words are put together in authentic text and is grammatically well-structured (Sinclair, 1991; Kjellmer, 1987). According to Benson, Benson, and Ilson (1986: 171), collocations fall into two major groups: lexical collocations and grammatical collocations. The difference between these two groups lies in the types of words involved. Lexical collocations roughly consist of syntagmatic affinities among open class words such as *verbs, nouns, adjectives, and adverbs.* In contrast, grammatical collocations generally involve at least one closed class word among *particles, prepositions,* and *auxiliary verbs.* Grammatical collocations are very similar to lexical collocations in the sense that they also correspond to arbitrary and recurrent word co-occurrences (Benson, 1990; Bahns and Eldaw, 1993). In short, the collocations are as the tendency of certain words to co-occur regularly in a language (Baker, 1992).

Language for Specific Purpose (LSP) like meteorology contains specialized collocations that are words which are regularly used together (Bowker and Pearson, 2002: 39). For example, in weather forecasts, we use 'strong wind' but 'heavy rain'. It would not be normal to use *heavy wind or *strong rain. The word 'strong' occurs with (collocate with) 'wind', and 'heavy' collocate with 'rain' (Runcie, 2002). Naturally put, collocations are regularly regarded as words which are frequently 'found in each other's company'. The identification of forms of word *co-occurrence* or specialized collocations in textual data is particularly important in dictionary making, natural language processing and language teaching.

The collocation knowledge is necessary for language learners if they want to produce native-like languages (Aroonmanakun, 2005). A competence in a language involves knowledge about collocation. English collocation exists and being able to use correct collocations greatly

helps a learner to master the language. The collocation allows us to think more quickly and communicate more efficiently and when learning a new word, it is important to learn it with its frequent co-occurrences or word partners, or a collocation. The learners can read quickly because they are constantly recognizing multi-word units rather than processing everything word-by-word. The collocation makes thinking easier. The collocation allows learners to name complex ideas quickly so that they can continue to manipulate the ideas without using all out brain space to focus on the form of words (Boonyasaquan, 2005; Lewis, 2000; James, 1998: 152; and Herbst, 1996: 389).

2.5 Abbreviations

Abbreviation is a reduction of a full word or phrase. It is a word or a phrase that expresses a shortened form. Abbreviations are divided into acronyms, initialisms, clipping, apheresises, contractions, substitutions, and symbols (Barnhart, 1995).

2.5.1 Initialism

Initialism is a word which is formed from the initials of a phrase and the word is pronounced as the resulting sequence of alphabets, such as *WMO* 'World Meteorological Organization', *TMD* 'Thai Meteorological Department', *UTC* 'Universal Coordinated Time ', *GMT* 'Greenwich Mean Time', *AMS* 'American Meteorological Society ', and *CDMS* 'Climate Data Management System'.

2.5.2 Acronyms

Acronyms are the combination of words which are reduced to initials and then pronounced as though the initials were exclusively alphabets in a typical word. Acronyms refer to words form from the beginning alphabets of words, especially the name of organizations as the following examples:

UNESCO	United Nations Educational, Scientific and Cultural Organization
NASA	National Aeronautics and Space Administration
TESOL	Teaching of English to Speakers of Other Languages

2.5.3 Clipping

The term 'clipping' is defined as a word or phrase which cut off some part of a word or phrase and thrown away the remains whereas the original meaning is retained. Clipping forms

shortened words from longer words, but does *not* change the part of speech. The clipping may be put at the initial of the word such as Temp (Temperature), Max (Maximum), Min (Minimum), Nov (November), Lat (Latitude), Long (Longitude), and Precip (Precipitation).

2.5.4 Aphaeresis

An aphaeresis is an omission at the beginning of a word. 1st from 'first', 2nd from 'second', 3rd from 'third', 4th from 'fourth', `coon' from `raccoon', and `till' from `until' are examples of the aphaeresis.

2.5.5 Contractions

The mechanisms of contractions are largely rooted in written English and they produce only occasional forms in spoken English, e.g. 'Dr.' (doctor) and 'Agt' (agent). A few contractions have entered spoken English and become quite prominent, such as 'TV' (television), 'IV'(intravenous), and 'QT' (quite).

2.5.6 Substitutions

A substitution is one kind of the abbreviation which it usually contains in English forms. It often involves abbreviation forms of foreign words and expressions that are adopted in English contexts such as 'A.M.' from Latin 'ante meridiem' (before noon), 'P.M.' from Latin 'post meridiem' (between noon and midnight), 'e.g.' from Latin 'exempli gratia' (for example), and 'lb' from Latin 'libra' (pound).

2.5.7 Symbols

According to Barnhart (1995), the symbols are classified as letter symbols, phonetics, and chemical symbols. Symbols are used more in scientific and technical literature than in any other area. The letters are abbreviations for particular words, but they stand as symbols in relation to one another for a concept and for quantities once the formula is applied in a specific instance (Barnhart, 1995: xviii). In formula $\mathbf{E = mc^2}$ (where **E** stands for energy, **m** for mass, and **c** for constant speed of light) and **SOS** (for distress), phonetic symbols in the pronunciation systems of dictionaries e.g. **ð** (for th), and chemical symbols in every periodic table of the elements e.g. **Au** (for gold) and **Ag** (for silver).

2.6 Vocabulary

Each vocabulary can be used in different situations. According to Nation (2001), vocabulary is divided into four levels: high frequency words; academic vocabulary; technical vocabulary; and low frequency words.

2.6.1 The General Service List (GSL)

Vocabularies refer to the most high frequency words of English which are generated by Michael West (1953). These words are called the General Service List (GSL) which refers to general in English words. It contains the most beneficial 2000 English word families providing both semantic and frequency information drawn from a corpus of five million words created from general written. This list aims to use as a basis for teaching and studying of vocabulary for non-native speakers of English. The GSL use criteria include frequency and text coverage. The selection of vocabulary was selected on the basis of practical application. John Bauman and Brent Culligan created the latest version of GSL in 1995. This new list consists of 2,284 words and expresses in frequency order based on the Brown Corpus in order to solve the problems of original version. Moreover, the frequency order was ranked according to the number of appearance in the 1,000,000 words of the Brown Corpus.

2.6.2 The Academic Word Service List (AWL)

The second level of vocabulary has been called academic vocabulary (Martin, 1976), sub-technical vocabulary (Cowan, 1974) or semi-technical vocabulary (Farrell, 1990). This vocabulary is common to a wide range of academic fields but is not what is known as high frequency vocabulary and is not technical in that it is not typically associated with just one field. Usually, this vocabulary is called Academic Word Service List (AWL) (Coxhead, 2000).

The Coxhead's list excludes words in the General Service List (West, 1953) or most frequent 2000 words in the English language (Vongpumivitch et al., 2009). Frequency of use was the main criterion in Coxhead's compilation. For a word to be included in the AWL, it has to occur at least 100 times in the entire corpus, 10 times in each of the four main disciplines, and at least once in 15 out of 28 subject areas that are included in the corpus. The final Academic Word Service List (AWL) contains 570 word families and represents approximately 10% of the total tokens in Coxhead's corpus (Coxhead, 2000). The coverage of the AWL words in the four disciplines is 9.3% in Arts, 12% in Commerce, 9.4% in Law, and 9.1% in Science (Coxhead,

2000). Cobb and Horst (2004) hold the view that the knowledge of the AWL words as well as the knowledge of words in West's (1953) General Service List (GSL) is needful in comprehension an academic text in English. Coxhead and Nation (2001) stated that the AWL word cover around 8.5-10% of the running words in academic texts.

2.6.3 Technical Vocabulary

The third level of vocabulary, technical vocabulary is a word closely related to a specific area or field of study, such as medicine, engineering, earth science, geology, geography, physics, statistic, linguistics, meteorology, and so on. Sometimes it is called specialized vocabulary, terminology or specific vocabulary. Nation (2001) defines the technical vocabularies as the words in specialized texts, and is made up of words that occurred frequently in a specialized text or subject area but did not occur or were of very low frequency in other fields. One of the prominent aspects of the linguistic knowledge needed to comprehend specialized texts is the corresponding specialized vocabulary or "technical words that are recognizably specific to a particular topic, field, or discipline" (Nation, 2001: 198). This type of vocabulary refers to a word or a group of words in the form of a noun phrase used as a technical lexical unit in a specific field with specific meaning to the field (Nation, 2001). It varies according to different fields of study.

According to Wasuntarasophit (2008: 30), the technical vocabulary can be a common word which provides a specialized meaning different from its vernacular meaning when used in specific field. He explained that researchers in the technical field have provided similar definitions for technical vocabulary regarding:

- Groups of specialists;
- Field of use;
- Special meaning;
- Distinct forms or morphology;
- High frequency of occurrence in the specific field.

According to Bowker and Pearson (2002: 25-26), people who study a foreign language also begin by learning the Language for General Purposes (LGP) of that language. Being conversant in the Language for General Purposes (LGP) of a language allows students to function in that language by doing things such as asking for directions, ordering a meal in a restaurant, chatting with a friend about the film you saw last night, or talking about the weather. All of these subjects can be discussed in a very general way using Language for General Purposes (LGP), and

they can also be discussed at a different, more specialized level using LSP. Fields of knowledge such as orienteering, gastrononym, cinematolography and meteorology are discussed by experts using the LSP appropriate to these fields. As an Language for General Purposes (LGP) speaker, you might feel a little out of your element if you overheard two meteorologist discussing the weather using terms like 'advection', 'helicity' and 'radiational cooling'. Here are some examples of specialized vocabulary from Bowker and Pearson (2002) as shown in Figure 2.2 below:

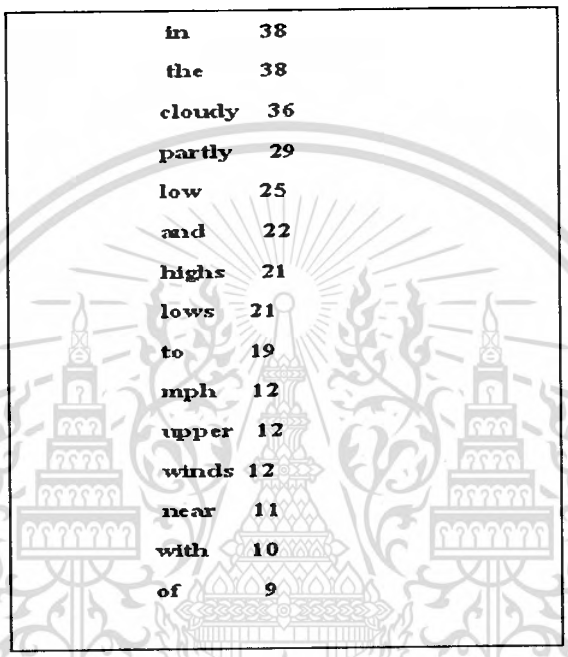


Figure 2.2 The Top Fifthteen Word Frequency Lists in a Small Corpus on Weather Forecasts

Figure 2.2 contains a word frequency list from a small corpus of weather forecasts. As you can see, some of the most common words in this corpus are general language words, such as 'in', 'the', 'and', 'to', 'with', 'of'; however, there are also some terms that are more specific to Language for Specialized Purposes (LSP) of weather forecasts, including 'cloudy', 'highs', 'low' and 'winds'. A corpus can identify some potentially interesting terms belonging to the Language for Specialized Purposes (LSP) by looking at these terms in context using a concordancer. The terms in context will help us to learn more about their meaning and behavior.

In Thailand, the meteorological field, terminology and creation of modern reference is lagged behind causing problems to translators and affecting their translations. Therefore, there is a need to study the vocabulary in weather forecast English. In this study, words are divided into three groups 1) general vocabulary, 2) academic vocabulary, and 3) technical vocabulary. The

separation of the general vocabulary is by using the General Service List (GSL) (Bauman and Culligan, 1995), and academic vocabulary using the Academic Word List (AWL) (Coxhead, 2000). Weather Glossary, the definitions of meteorology, and Glossary of meteorology and the meaning in the context are used to classify the technical vocabulary.

2.6.4 Miscellaneous

The fourth level of vocabulary consists of all the remaining words of English outside the General Service List (GSL), the Academic Word List (AWL), and the technical vocabulary. They include proper names, abbreviations, and numbers. They typically cover around 5% of the running words in texts.

2.7 Corpus Lexicography and Technical Dictionaries

2.7.1 The Definition of Corpus

Nowadays the terms 'corpus' (plural: corpora) refer to a collection of text, of the written or spoken word, which is stored and processed on computer for the purposes of linguistic research (Renouf, 1987: 1; Hartmann and James, 1998) and searched on a computer, and the frequencies of words in that corpus can be determined and listed by a fairly trivial computer program. Leech (1997: 3) defines corpus as the follows:

> *"A corpus is, of itself, a rich resource of authentic data containing structures, patterns and predictable features that are waiting to be "unlocked" by the human intelligence."*
>
> *Leech (1997: 3)*

Sinclair (1991: 171) also indicated that a corpus typically contains many millions of words: this is because it is recognized that the creativity of natural language leads to such immense variety of expression that it is difficult to isolate the recurrent patterns or the lexical structure of the language.

Actually, corpus is a systematic collection of texts which documents the usage features of a language or language variety. The practical use of computer for data-processing and the theorical advances of corpus linguistics have given lexicographers powerful tools for storage and retrieval of (written and spoken) data to describe all aspect of language, especially vocabulary, and to present the results in dictionaries. For English, the British Nation Corpus, the Bank of

English and a number of specialized corpora are now available, as are the facilities of the internet, for collection project (Hartmann and James, 1998).

A corpus can be beneficial with learning about the linguistics features of English for Specific Purposes: identifying specialized terms, learning about words that 'go together' (collocations), learning about grammar, learning about style, learning about concepts (Bowker and Pearson, 2002). In addition, the corpus tells us about semantic, syntactic, contextual, stylistic, and statistical, and then showing the difference between common and typical, giving accurate statistic, storing and recalling all the information that has been input, providing a vast number of real examples, giving more objective evidence, being permanently accessible to all, showing even recent changes in the language such as new words and new combination of words, and having the authority of the actual language use of many expert speakers (Krishnamurthy, 2000).

2.7.2 Corpus Lexicography

Corpus lexicography is a rapidly developing field, new methods and applications are appearing all the time (Morgan and Rinvolucri, 2004). The Collins COBUILD English Language Dictionary was the first significant completed example of corpus lexicography in English (Collins COBUILD, 1987). The lexicographers worked mostly with a concordance of about 7.3 million words (Krishnamurthy, 1987: 637quoted in Hockey, 2002). Corpus-based lexicography is an effective task for building a dictionary for languages, which exhibit explicit word boundaries (Potipiti et al., 2000). Kennedy's (1998: 33) studies that at the beginning in the late 1960s a number of electronic corpora were compiled for specialized purposes, especially, for lexicographical projects of various kinds.

The first computer corpora were developed as far back as the early 1960s. However, it was the computer boom of the 1980s and 1990s that established the corpus as an indispensable tool for lexicography. Mallikamas (1999: 1) stressed that a corpus-based analysis provides in depth empirical information on the language. There are several advantages in using computer readable forms of corpora including searching, sorting, retrieving and calculating linguistic data with high speed and accuracy. Likewise, one of the greatest advantages of using a corpus in lexicography is that it is possible to extract illustrative examples of the meaning and usage of Lemma (Leech et al., 2001: x). Corpora can be very useful resources for pursing various research agenda. For instance, many lexicographers have found that they can create dictionaries more effectively by studying word usage in very large linguistic corpora (Meyer, 2002: 11).

2.7.2.1 Types of Corpora

A corpus is always designed for a particular purpose, and the type of corpus will depend on its purpose (Hunston, 2002: 14). Usually, corpora depend upon the goal of the study, and may include things like whether the data consists of general or specific language.

2.7.2.1.1 General Corpora

A general corpus is a corpus of texts of many types. It may include written or spoken language, or both, and may include texts produced in one country or many. It is unlikely to be representative of any particular 'whole', but will include as wide a spread of texts as possible. A general corpus is usually much larger than a specialised corpus. It may be used to produce reference materials for language learning or translation, and it is often used as a base line in comparison with more specialised corpora. It is also sometimes called a reference corpus (Hunston, 2002: 14-15). There are many well-known general corpora such as the Brown Corpus, the LOB Corpus, the COBUILD corpus and the British National Corpus (BNC).

2.7.2.1.2 Specific Corpora

Specific corpus is a kind of corpora that are gathered for a specific purpose, and they vary in size and composition according to their purpose. Specific corpora are not balanced (except within the scope of their given purpose). Fuertes and Pedro (2007: 219) holds the view that specialized corpora – i.e., genre- and topic-specific corpora designed to include only samples of language of a particular type, belonging to a particular genre, register, etc. – have been used to help with LSP teaching and learning, to identify specialized terms, to learn about word collocations, and concepts.

Hunston (2002: 15) describes that a specialized corpus is a corpus of texts of a particular type, such as newspaper editorials, textbooks, and academic article in a particular subject, lectures, casual conversations, and so on. It is used to investigate the linguistic features that characterize a particular type of text.

To create the corpus for language learning for a specific purpose, Bowker and Pearson (2002) presume that well-designed corpora which comprise a number ranging from ten thousands to several hundred thousand words in size have proved to be exceptionally useful.

In this study, the corpus of weather forecast is the specialized corpus because it is relevant to one particular discipline, i.e. weather forecasts collected from many sources.

2.7.3 Technical Dictionary

Technical dictionaries refer to the traditional technical dictionaries. According to Opitz's (1983: 163) explanation, technical dictionaries are devoted to the description of the technical language of a specialized subject or discipline. They are compiled on the basis of criteria provided by particular target groups and their professional or special-interest needs. They deal with the language of a technical area. There are many references about dictionaries in the field of meteorology as follows:

The Dictionary of Global Climate Change was compiled by Maunder in 1992, which covers climatological and meteorological aspects of climate change. It consists of a scientific/technical component.

The Dictionary of Weather was compiled by Dunlop in 2001. It contains about 2000 definitions of weather, forecasting, and climate terms. Each entry contains English definition(s) and illustrative example(s). This dictionary includes 300 new and revised entries. His study was based on the related fields of oceanography, hydrology, climatology, and biographical information on important people in the development of meteorology. His aim was to make the dictionary of weather that can be a reference for meteorology and geography students, at school or university, as well as amateur meteorologists, general readers, and all those obsessed with the weather.

The Glossary of Meteorology (English-Thai) was combined by Thai Meteorological Department in 1979. The glossary consists of many sections about meteorology including cloud, temperature, pressure, etc. The lists were compiled manually.

Geographical Dictionary (English-Thai) (2006) was compiled by the Royal Institute. It contains many sections about geography and meteorology, the definition was created by Thai meteorologist in Meteorological Department.

2.7.4 The Corpus Inputs to the Technical Dictionary

The corpus is a primary source of information about the way words behave. It forms the basis of the way words combine with each other (syntactically and collocationally). It also provides information about word frequencies, grammatical patterns, and collocations. It is the main source of the example sentences shown in the dictionary (Macmillan: Online, 2009). Some information in the dictionary includes signs and symbols, lists of abbreviations, foreign words and phrases which were found in the corpus.

2.7.5 An Entry Structure of the Sample Dictionary

A good bilingual dictionary usually give information about the meaning, pronunciation, word classes or part of speech (POS), word grammar, collocations, example sentences, and synonyms (Redman and Edward, 1997: 8).The bilingual dictionary is easier for understanding and it is also good for working in English as much as possible. Ellendersen (2007) pointed out that information on word classes, morphology, and syntax ought to be included in Language for Specific Purposes (LSP) dictionary. Therefore, the sample dictionary should consist of the following entries: the English headword, abbreviation, grammatical information, Thai pronunciation, Thai synonym, Thai definition, English synonym, example of usage, and illustration.

2.7.5.1 Headword

The first part of the entry of the sample of meteorological dictionary is the headword which was focused on the technical single headwords, compounds, and abbreviations with high frequency in a corpus. In English language, there are many technical multiword items which may be combined from general, academic or other vocabularies in the same or other group of vocabularies in a specialized corpus. Thus, technical multiword terms or compounds should not be ignored and should be in the sample of technical dictionary as well.

2.7.5.2 Pronunciation and Word Stress

There are two types of information that they are required to utter a word suitably, they are the sounds and stress. If speakers pronounce clearly and correctly, their audience interlocutor should be able to understand what they are trying to express easily (Garrigues, 1999). Pronunciation is the aspect of uttering by human. It is a crucial skill that will contribute to non-native English speakers achieving goals of communication and is necessary for bilingual dictionaries. In the sample of bilingual (English and Thai) dictionary, Thai pronunciation and stress should be provided for users to pronounce the words correctly using the International Phonetic Alphabet (IPA) as a guideline because it is used practically in standard language for monolingual dictionaries and using standard references as Thai pronunciation's guideline.

2.7.5.3 Grammatical Information

The third item should be contained in the sample of technical dictionary, it is part of speech (POS) or known as word classes. It is important that the technical dictionary should provide grammatical information because it is essential for the non-native English speakers. In the sample dictionary, the basic grammatical information is provided in each entry, i.e. 'n.' for noun, 'v.' for

verb, 'adj.' for adjective, and 'adv.' for adverb. Because some headwords have many meanings, therefore, identifying the grammatical information will help separate the meaning of headword.

2.7.5.4 Definition

The important entry in the sample of bilingual (English and Thai) technical dictionary is Thai definition. Thai definitions in the sample dictionary will provide better understanding and they were defined by using the data from the technical references such as the Online Glossary of Meteorological Terms (English-Thai) (2007) and the Glossary of Meteorological Terms (1979) written by Thai Meteorological Department.

2.7.5.5 Abbreviation

In terms of technical dictionaries, technical abbreviations are very important. Most styles of scientific texts use many abbreviations. The technical abbreviations are useful for learning, reading, and writing the texts. Therefore, the technical abbreviations should be contained in the sample dictionary.

2.7.5.6 Example of Usage

A good dictionary includes examples of use, especially for words, the use of which is not self-evident. Manually compiled dictionaries usually have examples of use as well, for example, very little effort is put for considering the representativeness and coverage of the examples (Boas, 2005: 446). Computational processing makes it possible to retrieve all kinds of examples. The examples from the corpus provide information on collocations and grammatical information in authentic text. Therefore, in the sample of technical dictionary have to provide a number of examples to show typical contexts that the word is likely to be used in texts which were drawn from the concordance of the specialized corpus.

2.7.5.7 Synonym

Synonym is one of two or more words corresponding in the same meaning. Many headwords have synonyms, which make dictionary's users able to choose appropriate words for writing. Corpora can shed light on the precise relations and subtle distinctions of use among members of a set of similar items, at first glance synonymous (Partington, Online: 2005). The bilingual element helps to encourage students to read terms in contexts, and to develop resources in English. The sample of bilingual (English and Thai) technical dictionary gives Thai synonyms for the headword as well as the English synonyms. The synonyms were selected from the words which are similar in meaning with technical single headwords in a corpus. These synonyms in the sample of technical dictionary were based on standard references both Thai and English

2.7.5.8 Illustration

Illustrations such as diagram, charts, are used where their inclusions will be of practical value. They will enhance the users' understanding of the terms. Pictorial illustration is at best a representative of the type of concept defined. Illustrations can be quite helpful for improving comprehension in difficult technical vocabularies. The primary purpose of illustrations is to depict unusual or unfamiliar concepts. Technical and scientific dictionaries often use pictorial illustrations as well as charts and tables to supplement their texts. Therefore, the sample of technical dictionary should provide illustrations.

CHAPTER 3

RESEARCH METHODOLOGY

This chapter explains the research methods used in this study. It is divided into six main parts: research framework for the analysis, data collection, corpus design, research instruments, data analysis, and the design of the sample meteorological dictionary.

3.1 Research Framework for the Analysis

This research was an empirical research conducted by using corpus-based approach to analyze vocabulary of the English in weather forecasts in Thai Meteorological Department. In order to reach the goals and objectives of this study, the procedures were carried out as follows:

1. The English weather forecast texts were selected as the primary input for the Weather Forecast Corpus, and corpus structure was designed according to criteria of corpus design consideration.

2. The contents of the weather forecasts were scanned and downloaded into the computer readable forms.

3. Errors and unknown language which was scanned, were checked, edited, and deleted before analyzing by the lexical analysis program.

4. The concordance software 'WordSmith Tools' was utilized to create word frequency lists, count the number of tokens and types, and calculate the types/tokens ratio.

5. The groups of vocabulary were classified into nouns, verbs, adjectives, and adverbs.

6. The abbreviations were identified.

7. The vocabularies from the frequency lists were divided into three groups, namely general, academic, and technical vocabulary.

8. The headwords of the sample of meteorological dictionary were selected from the most frequent words from the list of technical single vocabulary, technical compounds, and technical abbreviation.

9. A sample of meteorological dictionary was designed.

3.2 Data Collection

3.2.1 The English weather forecasts texts were collected as the sample in this study. They were selected from three main sources: 1) websites: Thai Meteorological Department's website; World Meteorological Organization's website; 2) newspapers: Bangkok Post; and 3) documents in working units at Thai Meteorological Department. The categories of English weather forecasts texts were classified by using the weather forecast text classification from Thai Meteorological Department (TMD, 2007: Online).

3.2.2 These weather forecast texts were divided into two main categories, i.e. Thailand weather forecasts (WF) and meteorological documents used in Thai Meteorological Department working unit (MD). Thailand weather forecasts (WF) were divided into seven sub-categories: Weather Advisory (WF1), Daily Forecast (WF2), Shipping Forecast (WF3), Weekly Forecast (WF4), Monthly Forecast (WF5), Sunrise-Sunset and Moonrise-Moonset Forecast (WF6), and Water Level Forecast (WF7). In the category of the meteorological documents (MD) were divided into six sub-categories: Weather News (MD1), Weather Report (MD2), Meteorological Journal (MD3), Meteorological Magazine (MD4), Meteorological Textbook and Manual (MD5), and Other Meteorological Documents (MD6).

3.2.3 Five hundred fifty five thousand eight hundred and eighteen words of the weather forecast texts were randomly selected by using proportional stratified random sampling procedure. The proportional stratified random sampling was used because it is the basic sampling technique that every case has an equal chance of being selected for the sample (Healey, 1993). The size of the sample in each stratum was taken in proportion to the size of the stratum. The following formula was used to find the samples of this study (National Statistical Office, 2004: Online).

$$n_h = \frac{N_h}{N} \cdot n$$

n_h = Number of required-samples at stratum h

N_h = Total number of population at stratum h

N = Total number of all population

n = Total number of all required-samples

The samples in this study consisted of 55.6% in the Thailand weather forecasts (WF) stratum and 44.4% in the meteorological documents (MD) stratum. The relative sizes of the two main samples were 309,046 words of the WF and 205,046 words of the MD. This proportion was shown in Table 3.1 below:

Table 3.1 The Number of Weather Forecast Texts in the Weather Forecast Corpus (WFC)

Categories of Weather Forecast Texts	**Sub-categories of Weather Forecast Texts (2003-2009)**	**Number of Words in the Sub-categories**	**Percentage of Words in the Categories**
1.Weather Forecasts	1.1 Warning Advisory (WF1)	10,608	1.91%
	1.2 Daily Forecast (WF2)	63,686	11.46%
	1.3 Shipping Forecast (WF3)	53,392	9.61%
	1.4 Weekly Forecast (WF4)	54,561	9.82%
	1.5 Monthly Forecast (WF5)	23,104	4.16%
	1.6 Sunrise-Sunset and Moonrise-Moonset Forecast (WF6)	69,600	12.52%
	1.7 Water Level Forecast (WF7)	34,095	6.13%
		309,046	**55.6%**
2. Meteorological Documents	2.1 Weather News (MD1)	50,971	9.17%
	2.2 Weather Report (MD2)	25,130	4.52%
	2.3 Meteorological Journal (MD3)	4,636	8.34%
	2.4 Meteorological Magazine (MD4)	25,657	4.62%
	2.5 Meteorological Textbook and Manual (MD5)	76,968	13.85%
	2.6 Other Meteorological Documents (MD6)	21,684	3.90%
		205,046	**44.4%**
Total	**13 Sub-categories**	**555,818**	**100%**

3.2.5 Corpus Compilation

The samples of weather forecast texts were chosen from the printed weather forecasts which were scanned by using Optical Character Recognition (OCR), while electronic weather forecast texts were downloaded. The data obtained were stored as plain-text files (*.txt).

3.3 Corpus Design

Corpus design is the most important factor of the results that can be gained from such a study. Sinclair (1991: 13) describes that many people assume incorrectly that only a certain number of tokens are important for performing corpus analysis. In contrast, some people have used a very large corpus with a lack of consideration for the kind of text, and other people have done corpus design carefully but with a very small size. These two cases might greatly distort the research finding; therefore, corpus-based analysis researchers have to consider the methodology of corpus design and its creation.

This section is divided into four main parts: corpus tool, text-types for the corpus, time period, and corpus size.

3.3.1 Corpus Tool

In order to examine the most frequently occurring words in context and to reveal their most significant usage patterns, computer software, WordSmith Version 3.0 (Scott, 1998) was used to analyze the frequency of words and helped in the text analysis. The Word List, Concord, and Key Word are essential tools in this program. The concord was used to look up specialized words in the texts in this study. The ability of the concordance is that it sees many of examples of a word or phrase.

3.3.2 Text-types for the Corpus

3.3.2.1 Classification of Texts into Sub-domains

Thai Meteorological Department classified the weather forecasts into three categories as follows (TMD, 2007: Online):

1. Short Range Forecast (SRF) (not more than 72 hours)
2. Medium Range Forecast (MRF) (from 72 hours to 10 days)
3. Long Range Forecast (LRF) (from 10 days up)

In addition, this study consisted of two main categories of media, namely electronic, and both electronic and printed. Thus, these classifications were used in this study as shown in Table 3.2.

Table 3.2 Categories and Sub-categories of Samples, Media, and Period of Texts in the Weather Forecast Corpus (WFC)

Categories of Weather Forecast Texts	**Sub-categories of Weather Forecast Texts (2003-2009)**	**Categories of Media**	**Periods of Texts**	**Number of Words in the Sub-categories**	**Percentage of Words in the Categories**
1. Weather Forecasts	1.1 Warning Advisory (WF1)	Electronic	SRF	10,608	1.91%
	1.2 Daily Forecast (WF2)		SRF	63,686	11.46%
	1.3 Shipping Forecast (WF3)		SRF	53,392	9.61%
	1.4 Weekly Forecast (WF4)		MRF	54,561	9.82%
	1.5 Monthly Forecast (WF5)		LRF	23,104	4.16%
	1.6 Sunrise-Sunset and Moonrise-Moonset Forecast (WF6)		LRF	69,600	12.52%
	1.7 Water Level Forecast (WF7)		LRF	34,095	6.13%
2. Meteorological Documents	2.1 Weather News (MD1)	Electronic and Printed	Miscellaneous	50,971	9.17%
	2.2 Weather Report (MD2)			25,130	4.52%
	2.3 Meteorological Journal (MD3)			4,636	8.34%
	2.4 Meteorological Magazine (MD4)			25,657	4.62%
	2.5 Meteorological Textbook and Manual (MD5)			76,968	13.85%
	2.6 Other Meteorological Documents (MD6)			21,684	3.90%
			Total	**555,818**	**100%**

3.3.3 Time Period

The date of publication in texts is a significant criterion for selecting texts. The Weather Forecast Corpus focused mainly on texts published from 1 January 2003 to 31 December 2009. The period was regarded long enough to be representative of the language used in weather forecast texts.

3.3.4 Corpus Size

Corpus size usually depends on the purpose of each corpus study (Biber et al., 1998: 248-249). Issues of size also relate to the number of texts from different categories, the number of samples from each text, and the number of words in each sample. Coxhead (2000: 216-217) holds the view that in order to get an appropriate corpus, the research has to collect a reasonable number of words since more language seems to provide more chances of the occurrences and more lexical items can be examined in depth.

According to British Nation Corpus (BNC), 'Big is beautiful', this statement also support Sinclair about a criteria of a corpus size. Sinclair (1991: 18) expresses the view that in most of the very large texts, about half of vocabulary can be found only once in the texts. Actually, the corpus needs to contain millions of words, especially when analyzing the collection of vocabulary. A size of about 100,000 is also acceptable in the case of a specific corpus (Kaewphanngam, 2002: 29).

Therefore, the Weather Forecast Corpus should contain 500,000 words because it is specialized corpus.

3.4 Research Instruments

The instruments used in the study are Concordance Software and Meteorological Glossaries. The instruments were classified as follows:

3.4.1 Concordance Software

Concordance software is lexical analysis software for learning and researching about language. According to Simon (1996: 190), concordance software is a primary tool in all corpus study, and enables a researcher to find all occurrence of a given words, a part of word, or combinations of words within the context in a corpus of texts. There are several concordance programs. Bowker and Pearson's (2002: 130) suggest some practical concordance programs like

Keyword in Context Finder (KWiCFinder); Longi Term; MultiConcord; Para Conc; and WordSmith Tools.

The Weather Forecast Corpus was automatically analyzed using WordSmith Tools (WST) Version 3.0 (Scott, 1998). It is the concordance software used for anyone who needs to study texts closely or analyze language in depth by Oxford University Press. This program can be used for making indexes and word lists, word frequencies, comparing different usages of a word, analyzing keywords, and finding phrases and idioms (word-clusters) of the whole corpus. All vocabulary can be presented in terms of percentage of word occurrence. It can examine tokens and word types and calculate type/token ratio. This program consists of three main tools: Wordlist; Concord; and Keyword. The Concord was used for searching for specialized words in the texts. The WST Version 3.0 (Scott, 1998) can make a concordance. The aim of the concordance is to be able to see lots of examples of a word, phrase, or idiom in their texts. The WST has an ability to generate word lists. The word frequency lists were automatically presented in either alphabetical order or frequency order as shown in Figure 3.1 and 3.2 below.

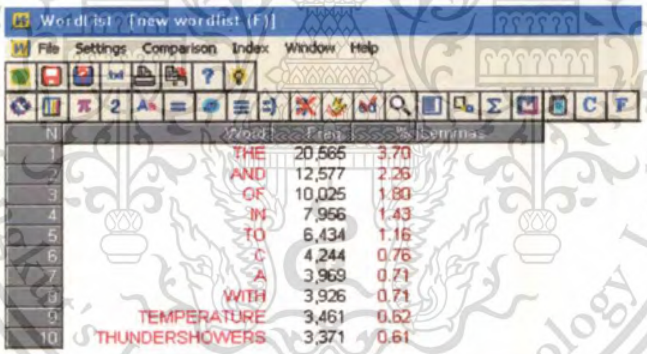


Figure 3.1 Frequency Lists

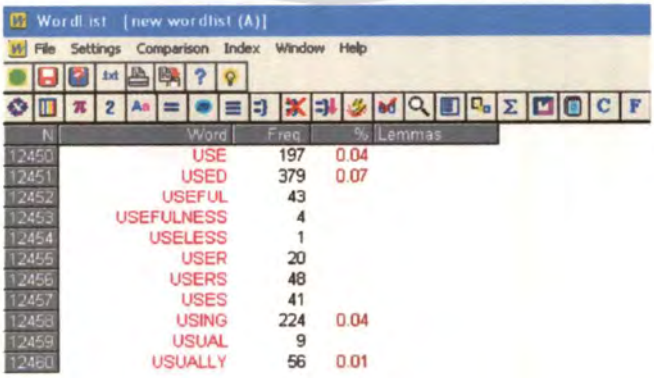


Figure 3.2 Alphabetical Lists

The Concord can present a concordance display, and give the researchers access to information about collocates of the search word. The purpose of a concordance is to be able to see lots of examples of a word or phrase in their contexts as shown in Figure 3.3.

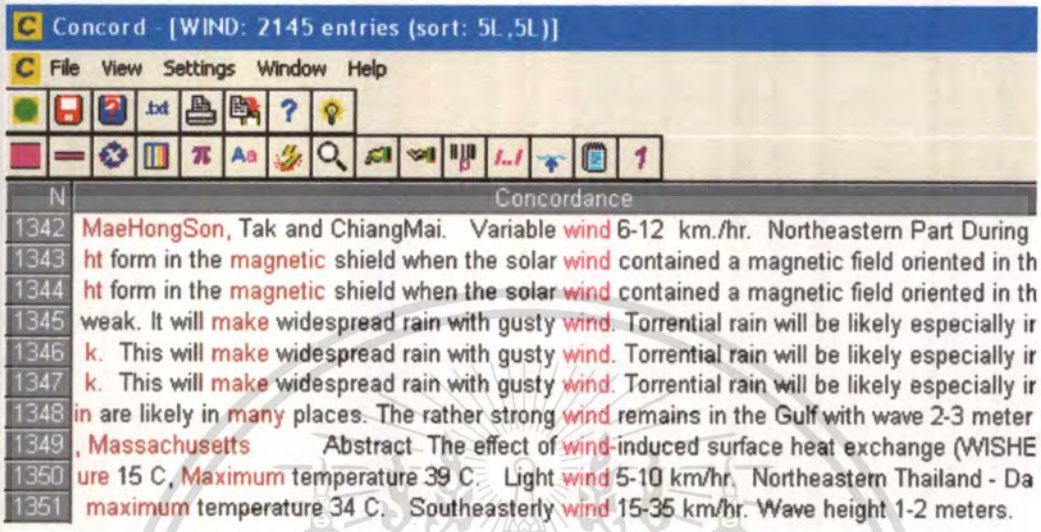


Figure 3.3 Concordance Lines of 'Wind'

Generally the Concord program is a tool for displaying repeated key words or phrases in context. These word clusters have the ability to show patterns of repeated phraseology or collocations in their concordance, particularly if they have a concordance with several thousand lines. The way through this program is press the icon '[C]', cluster in the Concord index. Bowker and Pearson (2002) stated that concordance program can produce lists of clusters, ranked alphabetically and by frequency. This is particularly important when working with LSP texts as the lexical unit is very often longer than a single word. The clustering facility enables us to identify multiword units. It will therefore help us with identifying units such as noun phrases which occur in a corpus. For example, 'maximum' and 'temperature' are word clusters in the noun phrase 'maximum temperature' (1,282) as shown in Figure 3.4.

N	cluster	Freq
1	temperature # c	2,243
2	maximum temperature #	1,282
3	minimum temperature #	1,264
4	temperature will drop	126
5	temperature at #	120
6	c # c	110
7	with morning fog	82
8	cool with morning	72
9	temperature will rise	71
10	maximum temperature at	66
11	average temperature #	53
12	minimum temperature at	53
13	cold in mountain	49
14	with maximum temperature	48
15	c rainfall mm	46
16	rainfall mm accumulative	46
17	max temperature #	44
18	the high mountain	44
19	the upper part	44
20	mm accumulative rainfall	41
21	temperature c rainfall	41
22	station temperature c	39

Figure 3.4 Clusters of 'Temperature'

Figure 3.4 shows the groups of noun phrases such as 'maximum temperature' (freq.1,282), 'minimum temperature' (freq.1,264), 'average temperature' (freq.53), 'max temperature' (freq.44), and 'station temperature' (freq.39). These clusters show not only noun phrases but also they show frequency of each a noun phrase of that searching word. In this case, the searching word is 'temperature' and noun phrases consist of 'temperature' are 'maximum temperature', 'minimum temperature', 'average temperature', 'max temperature', and 'station temperature'. It is clear that the cluster is highly useful for containing in a dictionary as well as single headwords.

3.4.2 Glossary of Meteorology

Glossaries of meteorology used consisted of 1) the Online Glossary of Meteorology written by American Meteorological Society (2000) which was used to define the technical vocabularies, 2) the Online Glossary of Meteorological Terms (English-Thai) (2007) and 3) Glossary of Meteorological Terms (1979) written by Thai Meteorological Department and used to define technical vocabularies. These references were developed to provide a source of authoritative, up-to-date information concerning weather forecast and meteorological terminology. Therefore, these references of meteorological terminology were chosen as tools for word selection, identification, definition of technical vocabulary in the sample of meteorological dictionary in this study.

3.5 Data Analysis

In this study, all weather forecasts texts were scanned and collected wholly as plain text files (.txt). There were 13 text files in the Weather Forecast Corpus, i.e. Warning News (WF1.txt), Daily Forecast (WF2.txt), Shipping Forecast (WF3.txt), Weekly Forecast (WF4.txt), Monthly Forecast (WF5.txt), Sunrise-Sunset and Moonrise-Moonset (WF6.txt), Water Forecast (WF7.txt), Weather News (MD1.txt), Weather Report (MD2.txt), Meteorological Journal (MD3.txt), Meteorological Magazine (MD4.txt), Meteorological Textbook and Manual (MD5.txt), and Other Meteorological Documents (MD6.txt). At the scanning procedure, the optical character recognition (OCR) software was used to accept scanned alphabets and represent the text electronically. All text files were sent out into Microsoft Word 2007 as word documents (*.doc) and the spelling of all words in the text was checked by using the spell-checker. The Wordlist Tool was used to compute the frequency of running words or tokens and word types, including type/token ratio. Finally, the researcher interpreted the results on word frequency counts and analysis of information collected.

3.6 The Design of a Sample of Meteorological Dictionary

The fourth objective in this study is to design a sample of meteorological dictionary based on the technical vocabularies found in the Weather Forecast Corpus. The design of the dictionary was based on 2.7.4 and 2.7.5 of Chapter 2. This section was divided into two main parts as follows:

3.6.1 The Corpus Inputs to the Technical Dictionary

The corpus is a primary source of information about the way words behave. It forms the basis of the way words combine with each other (syntactically and collocationally). It also provides information about word frequencies, grammatical patterns, and collocations. It is the main source of the example sentences shown in the dictionary (Macmillan: Online, 2009). Some information in the dictionary includes signs and symbols, lists of abbreviations, foreign words and phrases which were found in the corpus.

3.6.2 The Design of a Sample Meteorological Dictionary Entry

A good bilingual dictionary usually give information about the meaning, pronunciation, word classes or part of speech (POS), word grammar, collocations, example sentences, and

synonyms (Redman and Edward, 1997: 8). The bilingual dictionary is easier for understanding and it is also good for working in English as much as possible. Ellendersen (2007) pointed out that information on word classes, morphology, and syntax ought to be included in Language for Specific Purposes (LSP) dictionary. Therefore, the sample of meteorological dictionary should consist of the following entries: the English headword, abbreviation, grammatical information, Thai pronunciation, Thai synonym, Thai definition, English synonym, example of usage, and illustration as shown in Figure 3.5 below.

Figure 3.5 The Design of a Sample of Meteorological Dictionary Entry

3.6.2.1 Headword

The first part of the entry of the sample of meteorological dictionary is the headword which was focused on the technical single headwords, compounds, and abbreviations with high frequency in the Weather Forecast Corpus (WFC). In meteorological language, there are many technical multi-word items. Thus, technical multi-headwords should not be ignored and should be in the sample of meteorological dictionary as well.

3.6.2.2 Pronunciation and Word Stress

There are two types of information that they are required to pronounce a word suitably, they are the sounds and stress. They are crucial for non-native English speakers and necessary for bilingual dictionaries. In the sample of meteorological dictionary, Thai pronunciation and stress are provided for users to pronounce the words correctly using the International Phonetic Alphabet (IPA) as a guideline for Thai pronunciation because it is used practically in standard language for monolingual dictionaries and (Vitayapirak and Pornchalermpong, 2005).

3.6.2.3 Grammatical Information

Part of speech (POS) or word classes, is important for the non-native English speakers. Therefore, the technical dictionary should also provide grammatical information In the sample dictionary, the basic grammatical information is provided in each entry, i.e. 'n.' for noun, 'v.' for verb, 'adj.' for adjective, and 'adv.' for adverb. Some headwords have many meanings, therefore, identifying the grammatical information will help separate the meaning of headword.

3.6.2.4 Definition

Thai definitions in the sample dictionary will provide better understanding and they were defined by using the data from the Online Glossary of Meteorological Terms (English-Thai) (2007) and the Glossary of Meteorological Terms (1979) written by Thai Meteorological Department.

3.6.2.5 Abbreviation

Technical abbreviations are very important for weather forecast texts. The abbreviations in the Weather Forecast Corpus take 10.28% of text coverage or 30,784 tokens of the Weather Forecast Corpus. The abbreviations have high frequency in this study. Therefore, the technical abbreviations should be contained in the sample dictionary.

3.6.2.6 Example of Usage

A good dictionary includes examples of use (Boas, 2005: 446). Computational processing makes it possible to retrieve all kinds of examples. The examples from the corpus provide information on collocations and grammatical information in authentic text. The sample of meteorological dictionary must therefore provide a number of examples to show typical contexts that the word is likely to be used in weather forecast texts which were drawn from the concordance of the Weather Forecast Corpus.

3.6.2.7 Synonym

Synonym is one of two or more words corresponding in the same meaning (Partington, Online: 2005). The sample of meteorological dictionary gives Thai synonyms for the headword as well as the English synonyms. They were selected from the words which are similar in meaning with technical headwords in the Weather Forecast Corpus. The English synonyms in the sample of meteorological dictionary were based on the Online Glossary of Meteorology (AMS, 2000). The Thai synonyms in the sample of meteorological dictionary were based on the Online Glossary of Meteorological Terms (2007) and the Glossary of Meteorological Terms (1979) which were written by Thai Meteorological Department.

3.6.2.8 Illustration

Illustrations such as diagram, charts, are used where their inclusions will be of practical value. They will enhance the users' understanding of the terms. Pictorial illustration is at best a representative of the type of concept defined. Illustrations can be quite helpful for improving comprehension in difficult technical vocabularies. The primary purpose of illustrations is to depict unusual or unfamiliar concepts. Technical and scientific dictionaries often use pictorial illustrations as well as charts and tables to supplement their texts. Therefore, the sample of meteorological dictionary should provide illustrations.

CHAPTER 4

RESULTS AND DISCUSSIONS

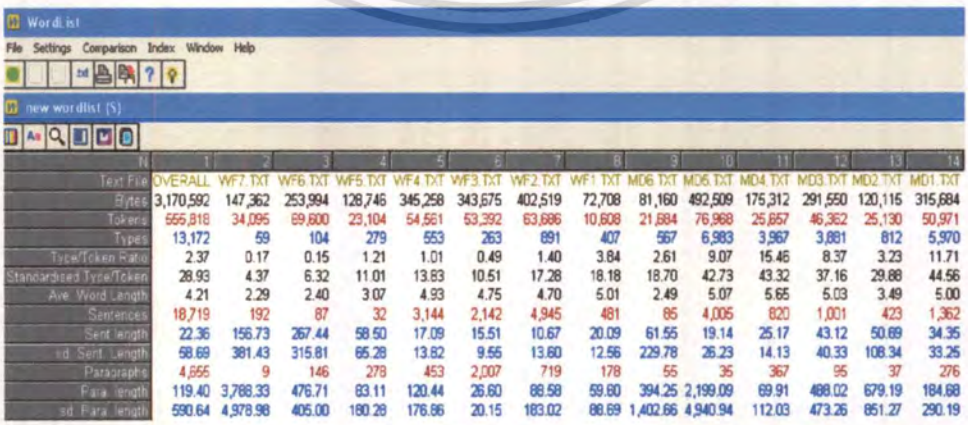
This chapter presents the results and discussions of the study followed by data analysis based on the objectives in Chapter 1. It is divided into eight main parts. The first part contains the statistical analysis and word frequency lists of the Weather Forecast Corpus (WFC). The second part analyzes the vocabulary in the Weather Forecast Corpus (WFC) into general, academic, and technical vocabulary. The third part focuses on the word classes in the corpus. The fourth presents the abbreviations in the study. The fifth part shows the collocations. The sixth part describes the compounds in this study. The seventh part displays the sample entries of a sample of meteorological dictionary. The last part in this chapter is the discussions on the results of this study.

4.1 Corpus Findings

This part is divided into Statistical Analysis and the Word Frequency Lists in the Weather Forecast Corpus (WFC).

4.1.1 Statistical Analysis of the Weather Forecast Corpus

The statistical details of the Weather Forecast Corpus display a summary of the content of a corpus. Figure 4.1 summarizes the number of tokens, word types and types/token ratio. Types/token ratio is computed by dividing the number of tokens by the number of word types.


WordList

File Settings Comparison Index Window Help

new wordlist (5)

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text File	OVERALL	WF7.TXT	WF6.TXT	WF5.TXT	WF4.TXT	WF3.TXT	WF2.TXT	WF1.TXT	MD6.TXT	MD5.TXT	MD4.TXT	MD3.TXT	MD2.TXT	MD1.TXT
Bytes	3,170,592	147,362	253,994	128,746	345,258	343,675	402,519	72,708	81,160	492,509	175,312	291,550	120,115	315,684
Tokens	555,818	34,095	69,600	23,104	54,561	53,392	63,686	10,608	21,684	76,968	25,857	46,362	25,130	50,971
Types	13,172	59	104	279	553	263	891	407	567	6,983	3,967	3,881	812	5,970
Type/Token Ratio	2.37	0.17	0.15	1.21	1.01	0.49	1.40	3.84	2.61	9.07	15.46	8.37	3.23	11.71
Standardised Type/Token	28.93	4.37	6.32	11.01	13.83	10.51	17.28	18.18	18.70	42.73	43.32	37.16	29.88	44.56
Ave. Word Length	4.21	2.29	2.40	3.07	4.93	4.75	4.70	5.01	2.49	5.07	5.65	5.03	3.49	5.00
Sentences	18,719	192	87	32	3,144	2,142	4,945	481	85	4,005	820	1,001	423	1,362
Sent. length	22.36	156.73	267.44	58.50	17.09	15.51	10.67	20.09	61.55	19.14	25.17	43.12	50.69	34.35
sd. Sent. Length	58.69	381.43	315.81	65.28	13.82	9.55	13.60	12.56	229.78	26.23	14.13	40.33	108.34	33.25
Paragraphs	4,655	9	146	278	453	2,007	719	178	55	35	367	95	37	276
Para. length	119.40	3,788.33	476.71	83.11	120.44	26.60	88.58	59.60	394.25	2,199.09	69.91	488.02	679.19	184.68
sd. Para. length	590.64	4,978.98	405.00	180.28	176.86	20.15	183.02	88.69	1,402.66	4,940.94	112.03	473.26	851.27	290.19

Figure 4.1 Statistical Details of the Whole Corpus

Figure 4.1 displays the Weather Forecast Corpus size which consists of 555,818 tokens or running words. The total number of word types equals 13,172 words in the corpus since a recurrent word is counted only once. The ratio of types/tokens was 1: 28.93. The ratio indicated that each word is repeated nearly 29 times on average throughout the corpus.

4.1.2 The Word Frequency Lists of the Weather Forecast Corpus

This section displays the word frequency list in the Weather Forecast Corpus. A word frequency list provides the number of time and percentage of each word that occur in the corpus. The list begins with the most frequent item and continues down to the single occurrences. The section presents three main word frequency lists. The first part shows the thirty most frequent words of the Weather Forecast Corpus (WFC). The second part compares the top twenty-five frequent words of the two specialized corpora, i.e. the Weather Forecast Corpus (WFC) and Pharmaceutical Corpus (Pongnan, 2006) with the other two general corpora, i.e. the British Nation Corpus (BNC) and COBULID Corpus. The last part presents the top twenty nouns in the Weather Forecast Corpus (WFC) compared with general corpus, the British Nation Corpus (BNC).

The following figure shows the top thirty frequent words and abbreviations occurring in the Weather Forecast Corpus (see Appendix A for more information).

N	Word	Freq	%	Lemma
1	THE	20,565	3.70	
2	AND	12,577	2.26	
3	OF	10,025	1.80	
4	IN	7,956	1.43	
5	TO	6,434	1.16	
6	C	4,244	0.76	
7	A	3,969	0.71	
8	WITH	3,926	0.71	
9	TEMPERATURE	3,461	0.62	
10	THUNDERSHOWERS	3,371	0.61	
11	KM	3,119	0.56	
12	FOR	3,107	0.56	
13	IS	2,869	0.52	
14	HR	2,765	0.50	
15	THAILAND	2,724	0.49	
16	SCATTERED	2,702	0.49	
17	AT	2,439	0.44	
18	FROM	2,384	0.43	
19	WAVE	2,259	0.41	
20	OR	2,170	0.39	
21	WIND	2,145	0.39	
22	HEIGHT	2,135	0.38	
23	DURING	2,056	0.37	
24	WEATHER	2,011	0.36	
25	METERS	1,958	0.35	
26	ISOLATED	1,956	0.35	
27	ON	1,945	0.35	
28	WINDS	1,819	0.33	
29	FORECAST	1,746	0.31	
30	BE	1,734	0.31	

Figure 4.2 The Top Thirty Frequent Words and Abbreviations in the Weather Forecast Corpus

Figure 4.2 displays the top thirty frequent words and abbreviations in the Weather Forecast Corpus. The closed class words occur in the top ten ranks. The closed class words are 'the', 'and', 'of', 'in', 'to', 'a', and 'with'. Within the Weather Forecast Corpus, the ten most frequently occurring types account for 11.78% of the total list. Therefore, the highest frequency word in the Weather Forecast Corpus is an article 'the' which occurs 20,565 times (3.70%). The conjunction with the highest frequency is the word "and" (freq.12,577 or 2.26%). According to Jurafsky and Martin (2002: 315), the word "and" is the most frequent conjunction in most English corpora and placed at the high rank. The open class words, 'temperature' appeared the highest with 3,461 occurrences or 0.62%. The open class word from Figure 4.2, namely 'temperature', 'thundershowers', 'scattered', 'wave', 'wind', 'weather', 'height', 'meters', 'isolated', 'winds', 'forecast' are highly related to weather forecast as the subject matter of the corpus. The first abbreviation 'c' occurs in the rank 6th. Generally, in the meteorology 'c' mean Celsius, Calm but in the prediction of weather 'c' often used as Celsius more than Calm. With regard to verb to be in the Weather Forecast Corpus, it is found that 'is', 'be' is the top thirty words in the rank 13th and 30th. The word 'BE' can also be used as other functions, but less frequently. The word 'BE' is identified as a verb which occurred 1, 618 times in the Weather Forecast Corpus while 'BE' as an abbreviation (B.E. for Buddhist Era) occurred 116 times.

The next table presents the comparison of the top twenty-five frequent words in specialized corpora: Weather Forecast Corpus (WFC), and Pharmaceutical Corpus with most words in general corpora: the British Nation Corpus (BNC), and the COBULID Corpus as shown in Table 4.1 below.

Table 4.1 Comparison of the Top Twenty-five Frequent Words in the Specialized and General Corpora

Specialized Corpora			**General Corpora**	
Rank	**WFC**	**Pharmaceutical Corpus**	**BNC (written texts)**	**COBOUILD Corpus**
1.	the	the	the	the
2.	and	of	of	of
3.	of	in	and	and

Table 4.1 (Continued)

Specialized Corpora			General Corpora	
Rank	**WFC**	**Pharmaceutical Corpus**	**BNC (written texts)**	**COBOUILD Corpus**
4.	in	and	a	to
5.	to	to	in	a
6.	a	with	to	in
7.	with	be	is	that
8.	temperature	is	to	I
9.	thundershowers	or	was	it
10.	for	a	it	was
11.	is	patients	for	is
12.	Thailand	mg	that	he
13.	scattered	should	with	for
14.	at	for	he	you
15.	from	may	be	on
16.	wave	dose	on	with
17.	or	not	I	as
18.	wind	treatment	by	's
19.	height	as	's	be
20.	during	been	at	had
21.	weather	by	you	but
22.	meters	are	are	they
23.	isolated	have	had	at
24.	on	at	his	his
25.	winds	on	not	have

Table 4.1 presents the comparison of the top twenty-five frequent words in two general corpora and two specialized corpora. The top ten words in the Weather Forecast Corpus are closed class words, namely articles, conjunctions, prepositions, determiners, for example 'the',

'and', 'of', 'in', 'to', 'a', 'with', and 'for'. These results are in substantial agreement with the study of Flowerdew (1993) that both general corpora and specialized corpora share similar aspect in that most ten frequent words. The closed class words are used in the texts of different fields.

However, in the specialized corpora such as 'temperature', 'thundershowers', 'Thailand', 'scattered', 'wave', 'wind', 'height', 'weather', 'meters', 'isolated', 'winds' in the ranks 8th, 9th, 12th, 13th, 16th, 18th, 19th, 21st, 22nd, 23rd, and 25th respectively and in the Pharmaceutical Corpus open class words are the words 'patients', 'dose', and 'treatment' in the ranks 11th, 16th and 18th respectively. Besides, it is also found that words which function as pronoun, i.e. 'I', 'you', 'it', 'he' and 'they' occur in the general corpora, but the twenty-five most frequent words in Table 4.1, they are not found in the specialized corpora, the Weather Forecast Corpus and Pharmaceutical Corpus. In two general corpora, the twenty-five most frequent words of the open class words are not contained in the general corpora. The word with highest frequency in both the general and specialized corpora is the word 'the'.

The next table illustrates the top twenty nouns in the Weather Forecast Corpus (WFC) compared with general, the British Nation Corpus (BNC).

Table 4.2 Comparison between the Top Twenty Nouns in the Weather Forecast Corpus (WFC) and British Nation Corpus (BNC)

Rank	Specialized Corpus: WFC		General Corpus: BNC (written texts)	
	Word	**Freq.**	**Word**	**Freq.**
1.	temperature	4,244	time	1819
2.	thundershowers	3,461	people	2063
3.	Thailand	3,371	years	704
4.	wave	3,119	way	1252
5.	wind	2,765	year	803
6.	height	2,724	government	295
7.	weather	2,259	work	601
8.	meters	2,145	man	405
9.	winds	2,135	world	248

Table 4.2 Continued)

Rank	Specialized Corpus: WFC		General Corpus: BNC (written texts)	
	Word	**Freq.**	**Word**	**Freq.**
10.	forecast	2,011	life	293
11.	area	1,958	day	746
12.	rain	1,819	part	355
13.	sea	1,746	house	460
14.	part	1,734	number	531
15.	gulf	1,723	children	367
16.	time	1,626	system	189
17.	areas	1,571	place	353
18.	knots	1,478	case	243
19.	Andaman	1,450	group	249
20.	period	1,357	company	158

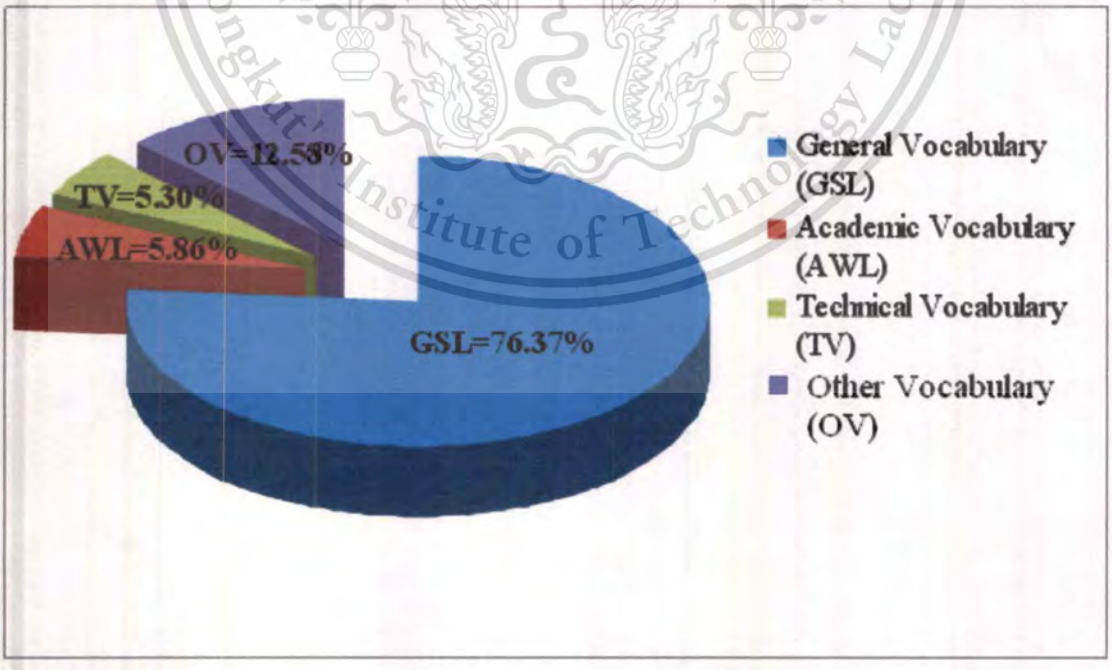
It can be noticed that in Table 4.2, the top twenty nouns in the BNC and WFC are different. For instance, the word 'temperature' appears first in the WFC but 'time' comes first in the BNC. Only one word occurs in both the corpora, namely, 'time' in the sixteenth rank in WFC and the first in BNC. From these differences, it is reasonable to accept the view of Sinclair (1991) that each corpus has its own emphasis according to the objectives, so the results of the corpus analysis clearly represent its characteristics. In fact, the results of each corpus analysis are concerned with the objectives; therefore, the different objectives and structures of a corpus design may lead to different findings. To conclude, the findings of this analysis reflect the lexical features of the WFC. It can be stated that the WFC is a specialized corpus which is certainly different from a general one like BNC. It is obvious that the Weather Forecast Corpus has its own distinctive lexicon such as 'temperature', 'thundershowers', 'wave', 'wind/winds', 'rain', 'knots', and so on.

4.2 Vocabulary in the Weather Forecast Corpus (WFC)

The second objective of this study is to analyze the vocabulary in the Weather Forecast Corpus by classifying them into three main types: general, academic, and technical vocabulary

from the total number of the word frequency list consisting of 555,818 tokens or 13,172 word types. However, the study of vocabulary in the Weather Forecast Corpus used only 268,600 tokens (593 lemmas) or 48.33% of the whole corpus which was selected from the samples in the study. The samples are repeated more than 55 times and have begun at 0.01% of the whole corpus because they are significances in the statistic. The classification of vocabularies into types was conducted manually after the general vocabularies, academic vocabularies, and technical vocabularies. All words in this study were divided into three types. The first type comprised general vocabularies that appeared in the General Service List (GSL) created by Bauman and Culligan (1995). The second type comprised the academic vocabularies that appeared in the Academic Word List (AWL) compiled by Coxhead (2000). The third type comprised of two groups of vocabularies, i.e. technical and other words outside those three groups. Technical vocabularies are identified by glossary of meteorology (AMS, 2000). The other words outside those three groups consist of proper names, technical vocabularies of other fields, etc. Then, the researcher checked every word from the lists again for the accuracy of the lists.

The 268,600 tokens were divided into three main types, namely general, academic, technical vocabulary and other kinds of words outside the three main types as displayed in Graph 4.1 below:



Graph 4.1 Proportions of Four Main Types of Vocabulary in the Weather Forecast Corpus

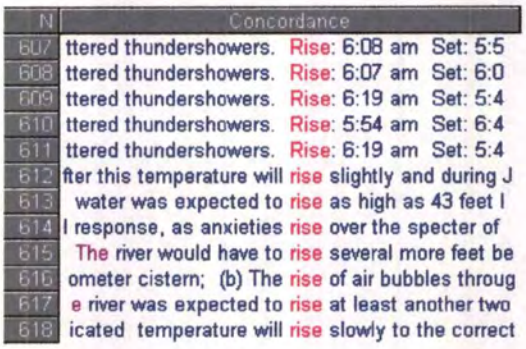
Graph 4.1 shows the proportions of different types of vocabulary which were used in this study in the Weather Forecast Corpus in percent of tokens (running words). Of the 593 lemmas or 268,600 tokens in the types of vocabulary used in this study, it is found that there are 354 lemmas (205,125) or 76.37% of general vocabulary of tokens, 51 lemmas (15,739) or 5.86% of academic vocabulary, 48 lemmas (14,252) or 5.30% of technical vocabulary, and 140 lemmas (33,792) or 12.58% of other vocabularies outside three groups such as 'sulphur', 'mercury', 'carbon', 'nitrogen', 'dioxide', 'Bangkok', 'Suratthani', 'Nakhonratchasima', 'Andaman', 'Pacific', 'Greenwich', and 'Alex'.

Among four types of vocabulary in this study, the general vocabulary has the highest proportion (76.37% of tokens), while the technical vocabulary has the lowest (5.30% of token). The academic vocabulary (5.86% tokens) is lower than the other vocabularies (12.58% of tokens). Some words from the general vocabulary, the academic vocabulary, and the other vocabularies can provide technical meanings in different contexts when they collocate with the vocabularies in three types of vocabulary in the Weather Forecast Corpus, for example, 'station temperature', 'extreme minimum temperature', 'sea surface temperature', 'soil temperature', 'heavy rain', 'widespread rain', 'almost widespread shower', 'torrential rain', 'low pressure cell', 'high pressure system', 'heat wave', and 'moderate sea'.

4.2.1 General Vocabulary in the Weather Forecast Corpus (WFC)

Of the 593 lemmas, the general words consisted of 354 lemmas (205,125 tokens) or 59.87% of general lemmas (76.37% of tokens). The high frequency of nouns in the Weather Forecast Corpus are 'temperature' (freq.3,461), 'wave' (freq.2,259), 'wind' (freq.2,145), 'height' (freq.2,135), and 'weather' (freq.2,011). Their meanings are related to weather forecasting in meteorology discipline. The findings of many nouns in the Weather Forecast Corpus showed that the meteorological language is informative by nature. The first general noun, adjective, verb, and adverb are 'temperature' (freq.3,461), 'heavy' (freq.1,617), 'rise' (freq.655), and 'very' (freq. 718) respectively.

In terms of the general verbs such as 'rise', 'set', 'change' appear in the present and future forms such as 'rise', 'set', 'change/changes', as shown in the following concordance data in Figure 4.3.



N	Concordance
607	ttered thundershowers. Rise: 6:08 am Set: 5:5
608	ttered thundershowers. Rise: 6:07 am Set: 6:0
609	ttered thundershowers. Rise: 6:19 am Set: 5:4
610	ttered thundershowers. Rise: 5:54 am Set: 6:4
611	ttered thundershowers. Rise: 6:19 am Set: 5:4
612	fter this temperature will rise slightly and during J
613	water was expected to rise as high as 43 feet I
614	l response, as anxieties rise over the specter of
615	The river would have to rise several more feet be
616	ometer cistern; (b) The rise of air bubbles throug
617	e river was expected to rise at least another two
618	icated temperature will rise slowly to the correct

Figure 4.3 Concordance Sample of *the* word 'rise' in the Weather Forecast Corpus

It can be seen from Figure 4.3 that the messages in the meteorological language in weather forecasting are often written in the form of report of daily situations and future situations of weather. Therefore, most of the full verbs in the Weather Forecast Corpus were used in the forms of the present simple and future simple tenses. The verb 'rise' used with adverb, modal verb, or preposition such as 'will rise slightly', 'will rise slowly', and 'rise over' was identified by the concordance. The overlap between verbs and nouns can be identified by using the concordance data from the corpus. For example, the word 'rise' (freq.655) has more than one function which can be nouns or verbs. For the study, the word 'rise' occurred 655 times which was used as a verb (freq.627 times) more than as a noun (freq.28).

4.2.2 Academic Vocabulary in the Weather Forecast Corpus (WFC)

Of the 593 lemmas, there are 51 lemmas (15,739 tokens) or 8.60% of academic lemmas (5.86% tokens). The academic adverbs are rarely found. The academic words such as 'area', 'minimum', 'issued', 'so' are most frequent open class words found in the type of academic vocabulary of the Weather Forecast Corpus.

4.2.3 Technical Vocabulary in the Weather Forecast Corpus (WFC)

Of the 593 lemmas, there are 48 lemmas (14,252 tokens) or 8.09% of technical lemmas (5.30% of token). The most frequent words in the type of technical vocabulary are nouns and adjectives. Verbs and adverbs in the technical vocabulary were not found in this study.

4.2.4 Other Vocabularies in the Weather Forecast Corpus (WFC)

Of the 593 lemmas, there are 140 lemmas (33,792 tokens) or 23.61% of texts coverage (12.58% of tokens). The most frequent words in the type of other words are proper names in this

study. There are 16,879 tokens or 69 lemmas (3.02% of texts coverage). The names of examples are displayed as follows:

1) countries: Thailand, China, Singapore, Mexico, Vietnam

2) nationalities: American, Thai

3) provinces in Thailand: Bangkok, Suratthani, Chumphon, Chiangmai, Phuket, Nakhonratchasima, Ubonratchathani, Songkhla, Tak, Chanthaburi, Ranong, Uthaithani, Nakhonsithammarat, Khon Kaen, Nakhonphanom, Kanchanaburi, Phangnga, Trat, Nan, Chiangrai, Nongkhai, Maehongson, Mukdahan, Phetchabun, Surin, Sakonnakhon, Phrae, Trang, Satun, Loei, Buriram

4) cities: Greenwich, Malacca, Kotabharu, Pattaya

5) states: Texas, Oklahoma

6) geographical places: Andaman, Atlantic, Pacific, Ko Samui

7) hurricane: Alex

Also, there are technical terms in other field (chemistry) such as 'sulphur' (chemical element), 'mercury' (chemical element), 'carbon' (chemical element), 'nitrogen' (chemical element), 'dioxide' (chemical compound), and 'monoxide' (chemical compound). Interestingly, the word 'dioxide'(freq.192) and 'monoxide'(freq.89) are chemical compounds which often occur with the word 'carbon' and 'nitrogen' and 'sulphur' in the Weather Forecast Corpus. These technical terms of the chemistry are often use in daily weather forecast in newspapers, Bangkok Post which deal with air quality indexes: ambient and roadside air quality indexes as shown in Figure 4.4 below.

Figure 4.4 Daily Weather Forecast of Thailand from Bangkok Post Newspaper

4. 3 Word Classes

According to Curzan and Adams (2006), word classes can be divided into two major classes: open class and closed class categories. There are both open class categories and closed class categories found in this study. They were presented as follows:

4.3.1 Open Class Categories

Open class categories consist of the six major closed class categories such as nouns, adjectives, verbs, and adjectives. The next table shows the top twenty of each major open class words in the Weather Forecast Corpus.

Table 4.3 The Top Ten of each Major Open Class Words in the Weather Forecast Corpus

No.	Nouns		Adjectives		Full Verbs		Adverbs	
	Word	**Freq.**	**Word**	**Freq.**	**Word**	**Freq.**	**Word**	**Freq.**
1	temperature	3,461	scattered	2,702	rise	655	widely	901
2	thundershowers	3,371	isolated	1,956	set	626	mostly	779
3	Thailand	2,724	heavy	1,617	issued	443	very	718
4	wave	2,259	cloudy	1,605	used	379	partly	521
5	wind	2,145	minimum	1,357	change	334	almost	475
6	height	2,135	maximum	1,301	expected	328	also	361
7	weather	2,011	southwesterly	985	broadcast	323	likely	308
8	meters	1,958	upper	957	prevails	313	below	230
9	winds	1,819	high	891	proceed	290	so	177
10	forecast	1,746	northeasterly	880	said	264	only	175

As can be seen from the top ten of each major open class in terms of word frequencies, the highest major open class word used in the Weather Forecast Corpus is noun. Most nouns in the Weather Forecast Corpus were related to infrastructures in weather forecasting such as temperature, wave, wind, weather, area, rain, sea, time and period. For example, *'Pattaya Scattered thundershowers. Slight to moderate sea. Wave height about 2 Meters. Minimum temperature 26 °C (79 °F). Maximum temperature 32 °C (90 °F)'*. In terms of adjectives, the highest adjective used in the Weather Forecast Corpus is 'scattered' (2,702 times), namely most

adjectives was used the most in weather forecasts. Similarly, the most adverbs in the corpus were also used in weather forecasts. The highest adverb in the Weather Forecast Corpus is 'widely'.

In terms of verbs in open classes of this study, most verbs are related to what things do in weather forecasting and what happened to them, for instance, 'rise', 'set', 'change' and 'expected'. Based on the top ten of open class words in the Weather Forecast Corpus, most of the full verbs in the Weather Forecast Corpus were used in the passive voice, present, and future tenses. For example: 'Scattered thundershowers with isolated heavy fall ***are expected*** over Southern Thailand from Surat Thani southward.' and 'Northeastern Part-Morning fog, temperature ***will change*** slightly'.

4.3.2 Closed Class Categories

Closed class categories consists of the six major closed class categories: prepositions, conjunctions, pronouns, complementizers, determiners, and auxiliary verbs (primary verbs and modal verbs) as displayed in Table 4.4 below.

Table 4. 4 The Top Five each Closed Class Words in the Weather Forecast Corpus

Types of Major Closed Classes	Examples
Prepositions	of (10,025), in (7,956), to (6,434), with (3,926), for (3,107)
Pronouns	all (622), some (612), its (379), such (337), other (292)
Determiners	the (20,565), a (3,969), this (1,550), that (1,543), more (500)
Conjunctions	and (12,577), or (2,170), as (1,170), than (551), but (332)
Complementizers	show that (26), said that (5), thought that (3), given that (3), say that (2)
Auxiliary verbs	Primary verbs: is (2,869), be (1,618), are (1,614), was (870), have (469)
	Modal verbs: will (1,247), should (1,175), can (464), may (298), would (194)

Table 4.4 displays the top 5 of each closed classes found in the Weather Forecast Corpus. The top five prepositions are 'of', 'in', 'to', 'with', and 'for'. The examples from the concordance data (most frequent clusters) are as follows:

(1) The Gulf **of** Thailand (Area 1-and 2) – very cloudy with almost widespread thundershowers.

(2) Central Part Cool with fog in the morning **in** the upper area.

The pronouns are 'all', 'some', 'its', 'such', and 'other'. The examples from the concordance data are as follows:

(1) **All** ships in the Gulf should proceed with caution and small boats should keep ashore.

(2) Northern Thailand – Hot in **some** places and widely scattered thundershowers mostly in Chaing Mai.

The words 'the', 'a',' this', 'that', and 'more' are determiners found in the Weather Forecast Corpus in the top five. The examples from the concordance data are as follows:

(1) WEATHER TODAY Southwesterly winds from **the** Andaman Sea bring humidity and hot temperatures across the country.

(2) **A** low pressure trough lies across the upper most Gulf.

The conjunctions are 'and', 'or', 'as', 'than ', and 'but'. The examples from the concordance data are as follows:

(1) THAILAND WEATHER BANGKOK Widespread thundershowers **and** isolated heavy rain.

(2) There were extensive rains in tropical East Africa, and some 15,000 people died **as** a direct result of floods and ensuing disease.

The auxiliary verbs can be divided into two types: primary verbs and modal verbs.

The primary verbs are 'is', 'be', 'are', 'was', and 'have'. The examples of primary verbs from the concordance data are as follows:

(1) Torrential rain and heavy fall **is** likely in the areas.

(2) All forms of transport should **be** aware of accident due to poor visibilities and showers.

The modal verbs are 'will', 'should', 'can', 'may', and 'would'. The examples of modal verbs from the concordance data are as follows:

(1) The system **will** be strengthening by January 23, slightly drop in temperature.

(2) All ships **should** proceed with caution in the thundershowers areas.

It was noticed that modal verbs in the Weather Forecast Corpus occurred at high frequency were 'will' (freq.1,247) and 'should' (freq.1,175). The message in the meteorological language in weather forecasting is often written in the form of report, instructing and warning for the situation in the future time (Srisirirungrot, 1998). Srisirirungrot also studied about weather

forecast by studying the form of weather forecast news presentation and its strategies in presenting news about disastrous weather its role. She found that weather forecast news generally had particular aspects and still the same when disaster occurs. In addition, bad weather will be added in the weather forecast news. She concluded that the major role of the weather forecast news presented by mass media is to inform about the weather. Moreover, the weather forecast news takes more roles in the warning and instructing. For instance, 'Torrential rain **will** be likely especially in much of the Northeast and East', and 'All ships **should** proceed with caution in the thundershowers areas'.

Clearly, 'will' was used for reporting the situations of weather in the future, while 'should' was used in instructing and warning when bad weather will occur.

The complementizers were found in the Weather Forecast Corpus, but less frequently. The examples from the concordance data are as follows: 'The provision precipitation data available so far **show that** in Alameda de Cervera (Ciudad Real) 240 mm fell between 8 a.m. on 23 and 8 a.m. on 24 May 2009'.

4.4 Abbreviations

Abbreviations take 10.28% of text coverage in this study or 30,784 tokens of the Weather Forecast Corpus. The top twenty abbreviations in the corpus are as follows (see Appendix F for more information):

Table 4.5 The Top Twenty of Abbreviations in the Weather Forecast Corpus

No.	Rank	Abbrev.	Meaning	Freq.	Percentage
1	6	°C	Celsius	4,244	0.76
2	11	km	Kilometer	3,119	0.56
3	14	hr.	Hour	2,765	0.50
4	46	Precip	Precipitation	1,239	0.22
5	47	°F	Fahrenheit	1,212	0.22
6	63	P.M./p.m.	Post Meridiem	849	0.15
7	67	AZ/Az	Azimuth	790	0.14
8	70	*A.M./a.m.	Ante Meridiem	751	0.14
9	10	m	Meter/Metre	727	0.13

Table 4.5 (Continued)

No.	Rank	Abbrev.	Meaning	Freq.	Percentage
10	75	UTC	Universal Time Coordinated	727	0.13
11	81	h.m.	Mounting High	687	0.12
12	86	No/no.	Number	648	0.12
13	87	max	Maximum/Minute	637	0.11
14	107	th.	Thailand	533	0.10
15	114	's	Is, Has, Was	482	0.09
16	115	min	Minimum, Minute	481	0.09
17	125	H	Height	461	0.08
18	131	Sa	Saturday	430	0.08
19	132	N	North	422	0.08
20	138	Fr	Friday	412	0.07
			Total	**21,369**	**3.89**

*The abbreviations can also be used as other functions, but less frequently.

Table 4.5 presents the top twenty of abbreviations in the Weather Forecast Corpus. Interestingly, most abbreviations in this corpus were concerned with temperature (e.g. measurement units of temperature: °C (Celsius) and °F (Fahrenheit), types of temperature: max (Maximum temperature), min (Minimum temperature), time (e.g. hr. (Hour), P.M./p.m. (Post Meridiem), A.M./a.m. (Ante Meridiem), UTC (Universal Time Coordinated), min (Minute)), wind (e.g. hr. (Hour) and height of wave (e.g. m (Meter/Metre), H (Height)). Interestingly, the overlap between abbreviations and verbs can be identified by using the concordance data from the corpus. For example, the word 'am' (freq.752)' has more than one function which can be nouns (abbreviations) or verbs. In this study, it was found 751 times the abbreviations 'a.m.' (for Ante Meridiem) and 1 time as a verb. Therefore, the words 'am' in the Weather Forecast Corpus often were used as abbreviation 'Ante Meridiem'. Since the verb 'am' rather use in spoken texts but in the Weather Forecast Corpus is a written text. Examples of concordance lines 'am' are displayed in Figure 4.5.

740	n Wednesday: Minimum temperature at 6.20 am 20.6 C. 69.1 F. : Maximum temperature
741	n Wednesday: Minimum temperature at 1.15 am 25.0 C. 77.0 F. : Maximum temperature
742	n Wednesday: Minimum temperature at 5.00 am 25.3 C. 77.5 F. : Maximum temperature
743	NGKOK Cool with isolated rain. Rise: 6:38 am Set: 5:57 pm Rise: 12:58 pm Set: 1:00
744	KOK Morning cool with light fog. Rise: 6:33 am Set: 5:52 pm Rise: 5:19 pm Set: 4:49 a
745	KOK Morning cool with light fog. Rise: 5:53 am Set: 6:35 pm Rise: 5:01 pm Set: 4:09
746	KOK Morning cool with light fog. Rise: 6:24 am Set: 5:47 pm Rise: 3:07 pm Set: 3:12 a
747	NGKOK Cool with isolated rain. Rise: 6:33 am Set: 5:52 pm Rise: 6:13 am Set: 5:42
748	GKOK Cool with isolated rain. Rise: 6:38 am Set: 5:56 pm Rise: 12:19 pm Set: 12:1
749	d like the creek would flood again. "But where am I going to go? My wife is ill. We don't hav
750	a, Associated Press Writer – Wed Jul 7, 8:52 am ET NEW YORK – Hot town, summer in t
751	a, Associated Press Writer – Wed Jul 7, 8:52 am ET NEW YORK – Hot town, summer in t
752	t: 66 Wind: WSW 2 mph Sunrise: 6:45 am Sunset: 6:16 pm 90 High: 88 Low: 7

Figure 4.5 Concordance Sample of 'am' in the Weather Forecast Corpus

The abbreviation 'C' is the highest frequent abbreviation in the Weather Forecast Corpus. 'C' is a general abbreviation that appears in many areas. It is the mark of inductive reasoning as the writer uses it to support any generalization with the specific case. It can be seen from the concordance sample of Figure 4.6 that 'c' in the Weather Forecast Corpus (WFC) was used in a specific area (meteorology). It was often used with temperature, maximum temperature or minimum temperature.

part with minimum temperature 16-23 C. Hot to very hot during the d
Hot with maximum temperature 35-38 C. and widely scattered to scat
part with minimum temperature 13-18 C while cool elsewhere, Minimu
KOK Cool with morning fog. min. 21 C max. 32 C Rise: 6:35 am S
proceed with caution. North Min. 13 c Max. 32 c Cold and morning
h scattered thundershowers. min. 25 c max. 33 c Rise: 6:05 am S
y with cool. Minimum temperature 17 C. Maximum temperature 30 C.
part with minimum temperature 13-18 C while cool elsewhere, 18-23.

Figure 4.6 Concordance Sample of the Abbreviation 'C' in the Weather Forecast Corpus

The following figure shows the concordance for the technical abbreviation 'Precip' in the Weather Forecast Corpus.

Concordance				
Saturday	1	Actual: 31 | 21	Precip: 0.00	Average: 31 | 22
Saturday	1	Actual: 31 | 25	Precip: 0.00	Average: 33 | 26
Saturday	1	Actual: 31 | 25	Precip: 0.00	Average: 34 | 25
Saturday	1	Actual: 31 | 25	Precip: 0.00	Average: 34 | 26
y Saturday	1	Actual: 35 | 26	Precip: 0.00	Average: 34 | 25
Saturday	1	Actual: 32 | 25	Precip: 0.00	Average: 33 | 26
Saturday	1	Actual: 32 | 25	Precip: 0.00	Average: 32 | 24
Saturday	1	Actual: 35 | 27	Precip: 0.00	Average: 33 | 26
Saturday	1	Actual: 33 | 25	Precip: 0.00	Average: 32 | 26

Figure 4.7 Concordance Sample of the Abbreviation 'Precip' in the Weather Forecast Corpus

It can be seen from Figure 4.7 that the technical abbreviation 'Precip' (Precipitation) occurs in area of meteorology in the weather forecasts. It means all liquid or solid phase aqueous particles that originate in the atmosphere and fall to the earth's surface (AMS, 2000) including rain, snow and hail. The top four words of each type of abbreviations in the Weather Forecast Corpus are shown in Table 4.6 below.

Table 4.6 The Top Four of each Type of Abbreviations in the Weather Forecast Corpus

No.	Types	Examples of Abbreviations	Meaning	Freq.	%
1.	Clipping	Precip	Precipitation	1,239	0.22
		AZ/Az	Azimuth	790	0.14
		max	Maximum	637	0.11
		th.	Thailand	533	0.10
2.	Initialism	°C	Celsius	4,244	0.76
		km (s)	Kilometer(s)	3,119	0.56
		B.E.	Buddhist Era	1,734	0.31
		°F	Fahrenheit	1,212	0.22
3.	Contraction	hr.(s)	Hour(s)	2,765	0.50
		No/no.	Number	648	0.12
		et al.	And Others/And elsewhere	200	0.04
		avg	Average	149	0.03
4.	Acronym	NOAA	National Oceanic Atmosheric Administration	159	0.03
		NCEP	National Centres for Environmental Prediction	141	0.03
		TIGGE	THORPEX Interactive Grand Global Ensemble	79	0.01
		NASA	National Aeronautics and Space Agency	64	0.01

Table 4.6 (Continued)

No.	Types	Examples of Abbreviations	Meaning	Freq.	%
5.	Substitution	P.M./p.m.	Post Meridiem	849	0.15
		A.M./a.m.	Ante Meridiem	751	0.14
		e.g.	for example	93	0.02
		A.D.	Anno Domini	88	0.02
6.	Symbol		Rain	690	0.12
			Waves	610	0.11
			Temperature	540	0.10
			Partly Cloudy	440	0.08

4.5 Collocations

For the study of collocation and phraseology, it is necessary to study huge amounts of text in order to isolate the recurrent patterns of individual words which are necessary evidence with which to base the generalizations of grammar (Sinclair, 1991: 20). This study aims to analyze collocations in terms of nouns, adjectives, verbs, and adverbs. The concordance data displays combinations of words or collocations. The collocations of nouns, adjectives, verbs, and adverbs are shown in the next sections.

4.5.1 The Collocation of the Noun 'Temperature'

'Temperature' functions as noun in open classes. It is the highest frequently used noun occurring in the Weather Forecast Corpus with the frequency of 3,461 or 0.62%. The collocations of the word 'temperature' are shown in Figure 4.8.

Vicinity - Hot with widely scattered thundershowers. Minimum temperature 25 C, maximum temperature 36C. Southerly winds 15-30 km/hr. Today : F
t submitted to Mon. Wea. Rev.). With this approach, the virtual temperature used in the CFSR AM is replaced by specific enthalpy for all prognostic ec
Southern Part(west Coast) Cool with morning light fog, minimum temperature 20-24 C . Northeasterly wind 15-35 km./hr . Wave height 1-2 meters, offsh
ion analysis of the reservoir inflows with the original sea surface temperature anomaly data, the principal sources of variability in Sydney's reservoir inflow
during the day with widely scattered thundershowers, maximum temperature 34-36 C, except during March 28-30, scattered thundershowers are likely.
in media with nonhomogeneous thermodynamic properties (ie., temperature and pressure), containing a mixture of absorbing gases with variable concen
Metropolis and Vicinity Cool with morning light fog, minimum temperature 17-21 C. During February 4-8, temperature will rise. Northeasterly wind 15-
ions showed good agreement with selected observations of soil temperature. Major results show that, during the rainfall season, soil moisture in the first

Figure 4.8 Collocations of Noun 'temperature' in the Weather Forecast Corpus

As can be seen in Figure 4.8, the noun 'temperature' often co-occurs with adjectives and nouns such as minimum temperature occurring 1,284 times and maximum temperature occurring 1,136 times. The lexical collocations of 'temperature' showed what type of temperature occurred in the Weather Forecast Corpus. Table 4.8 presents the occurrences of lexical collocations (adjective + noun) and (noun + noun) of 'temperature', their frequency, and examples of uses.

4.5.2 The Collocation of Adjective 'Scattered'

'Scattered' functions as adjective in open classes. It is the highest frequent adjective occurring in the Weather Forecast Corpus with a frequency of 2,702 or 0.49% of the total lists. The collocations of the word 'scattered' are shown in Figure 4.9.

Figure 4.9 Collocations of Adjective 'scattered' in the Weather Forecast Corpus

As presented in Figure 4.9, the adjective 'scattered' regularly co-occurs with adverbs and nouns, e.g. 'thundershower/thundershowers' and 'rain' and can be used with adverb 'widely'. The word 'scattered' is used before nouns and after adverbs such as 'scattered thundershowers' occurring 2,048 times and 'widely scattered thundershowers' occurring 553 times. With these results, we can see from the lexical collocations of 'scattered' what kind of rain occurred in the weather forecast texts in the Weather Forecast Corpus. Table 4.7 shows the occurrences of lexical collocations (adverb + adjective + noun) and (adjective + noun) of 'scattered', their frequency, and examples of uses.

4.5.3 The Collocation of Verb 'Rise'

'Rise' functions as verb in open classes. It is the highest frequent verb occurring in the Weather Forecast Corpus with the frequency of 655 or 0.12%. The collocations of the word 'rise' are displayed in Figure 4.10.

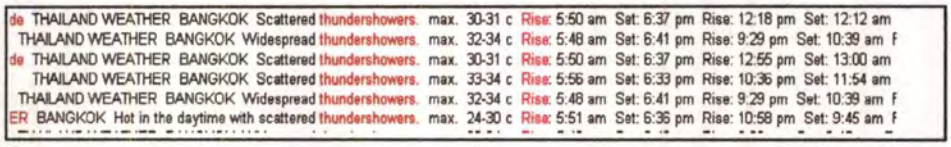


Figure 4.10 Collocations of Verb 'rise' in the Weather Forecast Corpus

Figure 4.10 displays the collocations of the verb 'rise'. Usually, most of meteorological language in weather forecasts are symbols. For example, 'THAILAND WEATHER, BANGKOK, Widespread thundershowers. Max. 32-34 °c Rise: 5:48 a.m. Set: 6:41 p.m. Rise: 9:29 p.m. Set: 10:39 a.m.'. The verb 'rise' shows time of sunrise and moonrise, there is not any subject, but displays the pictures of the sun and moon. The verb 'rise' was used as specific style in weather forecasts, namely, it can be seen that the word 'rise' from Figure 4.8 was used only in the singular, never in the plural. From the collocations of 'rise', it can be noticed that the style of weather forecasts is very sparse, sentences are very short, and there are no grammar words such as articles (e.g. 'the' or 'a'), used with nouns (e.g. 'sun or moon'). The information is presented in a systematic way as supported by Bowker and Pearson (2002). Usually, the short weather forecast usually begins with a weather forecasting area, a description of rain, moves on to temperature and ends with the time of sunrise, sunset, moonrise, and moonset.

4.5.4 The Collocation of Adverb 'Widely'

'Widely' functions as verb in open classes. It is the highest frequent verb occurring in this the Weather Forecast Corpus with the frequency of 901 or 0.16%. The collocations of the word 'widely' are displayed in Figure 4.11.

Figure 4.11 Collocations of Adverb 'widely' in the Weather Forecast Corpus

Figure 4.11 demonstrates the collocations of adverb 'widely', it often occurred before the

adjectives and nouns such as 'widely scattered thundershowers', 'widely scattered light rain', 'widely scattered thunderstorm', and 'widely scattered to scattered thundershowers'. As stated in Chapter 2, there are two types of collocations, i.e. lexical and grammatical collocations. The lexical collocations are phrases consisting of nouns, adjectives, verbs, or other elements. The grammatical collocations are phrases consisting of prepositions, conjunctions, pronouns, complementizers, determiners, and auxiliary verbs. The following two tables below show the frequent lexical and grammatical collocations in the Weather Forecast Corpus that were selected from the highest frequency of clusters.

Table 4.7 Lexical Collocations in the Weather Forecast Corpus

No.	Word	Lexical Collocations	Examples
1.	temperature (freq.3,461)	maximum temperature (freq.1,282)	Bangkok Metropolis and Vicinity Hot with day haze, ***maximum temperature*** 35-37 °C.
		minimum temperature (freq.1,264)	Eastern Part and Coastal Cool with fog in the morning, ***minimum temperature*** 21-23 °C.
		mean temperature (freq.72)	Monthly ***mean temperature*** was considered to be normal in November while it was slightly below normal in December.
		average temperature (freq. 53)	Minimum temperature at 6.30 a.m. 18.7 °C 65.7 °F: Maximum temperature at 2.30 p.m. 31.0 °C 87.8 °F: ***Average temperature*** 24.1 °C 75.4 °F: Average humidity 55 percent.
2.	thundershowers (freq. 3,371)	scattered thundershowers (freq.2,048)	The gulf of Thailand (Area1-and2)-Partly cloudy in the upper Gulf and cloudy with ***scattered thundershowers*** in the lower area.
		widely scattered thundershowers (freq.553)	Northern Thailand- ***Widely scattered thundershowers*** mostly in Mae Hong Son, Chiang Mai and Tak.

Table 4.7 (Continued)

No.	Word	Lexical Collocations	Examples
2.	thundershowers (freq. 3,371)	widespread thundershowers (freq.482)	Bangkok-***Widespread thundershowers***. Max. 32-34 °C Rise: 5:48 a.m. Set: 6:41 p.m. Rise: 9:29 p.m. Set: 10:39 a.m.
		isolated thundershowers (freq.468)	Central Thailand-Cool with morning fog and ***isolated thundershowers***.
		almost widespread thundershowers (freq.377)	Northeastern Part cloudy with scattered ***almost widespread thundershowers*** throughout of this period.
3.	scattered (freq.2,702)	widely scattered thundershowers (freq.553)	Northern Part ***Widely scattered thundershowers*** throughout of the period.
		scattered thundershowers (freq.2,048)	Central-***Scattered thundershowers***. Wind 15-30 kph.
		scattered thunderstorms (freq.34)	Northeastern Part S***cattered thundershowers*** with gust, isolated heavy rain and hail in some place.
		scattered rain (freq.30)	Weather Today ***Scattered rain*** with heavy rainfalls are likely in much of the lower North, the Central and the upper South.
		scattered showers (freq.27)	Forecast for Thailand Valid to Saturday morning. Chiang Mai Cloudy with ***scattered showers***.
4.	wave (freq.2,259)	wave height (freq.2,016)	A rather strong northeast wind prevails over the Gulf. ***Wave height*** about 2 meters has effect the Gulf and all ships should proceed with caution.

Table 4.7 (Continued)

No.	Word	Lexical Collocations	Examples
4.	wave (freq.2,259)	high wave (freq.61)	The prevailing strong northeast monsoon over the Gulf and southern Thailand causes ***high wave*** 2-4 meters in the Gulf of Thailand and monsoon surge.
		strong wind-wave (freq.17)	Rather ***strong wind-wave*** is likely in the Andaman Sea and the Gulf with wave height 2-3 meters.
		heat wave (freq.16)	More than 500 deaths in Hungary were attributed to the ***heat wave***, while major and widespread wildfire destroyed large areas of forest across the region.
5.	wind (freq.2,145)	northeasterly wind (freq.683)	***Northeasterly wind*** 11-22 knots or 20-40 km/hr. Wave height about 2 meters.
		easterly wind (freq.225)	***Easterly wind*** prevails over the Andaman Sea.
		gusty wind (freq.62)	Thundershower with ***gusty wind*** and heavy rain are expected in the areas during a few days.
		gusty wind (freq.62)	Thundershower with ***gusty wind*** and heavy rain are expected in the areas during a few days.
		southerly wind (freq.57)	Chumphon northward: ***southerly wind***, SuratThani southward: easterly winds 8-18 knots or 15-35 km/hr.
		wind shear (freq.30)	Increased ***wind shear*** in an Elnino tends to rip apart embryonic storms in the Atlantic basin and Caribbean Sea.
6.	height (freq.2,135)	wave height (freq.2,016)	***Wave height*** 1-2 meters, except for about 2 meters in the thundershowers areas.

Table 4.7 (Continued)

No.	Word	Lexical Collocations	Examples
6.	height (freq.2,135)	cloud height (freq.13)	The unit of measurement of ***cloud height*** is the metre or, for some aeronautical applications, the foot.
		geopotential height (freq.9)	The ***geopotential height*** is provided on the same surface plus 50 hPa.
		ceiling height (freq.8)	RUC20 and WRF-NMM forecasts of continuous ***ceiling height*** and visibility were interpolated to stations and converted into categorical forecasts.
		daily weather forecast (freq.371)	***Daily weather forecast Distribution*** On Monday 15 January 2007 Time issued 0600 LST General Situation High pressure system covers upper Thailand.
7.	weather (freq.2,011)	weather forecast (freq.147)	The ***weather forecast*** will be based on data collected from satellite images, radars, wind profilers and automatic weather stations.
		weather forecast bureau (freq.111)	The advisory is in effect for Thailand from 29 September 2009. Issued at 21.30 p.m. ***Weather Forecast Bureau***, Meteorological Department Ministry of Information and Communication Technology.
8.	isolated (1,956)	isolated heavy rain (freq.683)	Kotabharu to Singapore Route (Area4) – Very cloudy with almost widespread thundershowers and ***isolated heavy rain.*** Central Thailand – Morning light fog and ***isolated thundershowers*** in Kanchanaburi, Ratchaburi, Lop Buri, and Sara Buri.

Table 4.7 (Continued)

No.	Word	Lexical Collocations	Examples
8.	isolated (1,956)	isolated thundershowers (freq.468)	Northern Part Scattered to almost widespread thundershowers with ***isolated heavy fall***, after this period rain will decrease.
		isolated heavy fall (freq.115)	Southern Part (East Coast) Scattered thundershowers throughout the period, ***isolated heavy fall*** from NakhonSiThammarat southward during this period.
		isolated showers (freq.89)	Southern Part (East Coast) Hot in some areas and ***isolated showers*** mostly in the lower part.
		isolated light rain (freq.40)	Eastern Part and Coastal Morning cool in the upper part, and ***isolated light rain.***

Table 4.8 Grammatical Collocations in the Weather Forecast Corpus

No.	Word	Grammatical Collocations	Examples
1.	of (freq.10,025)	beware of (235)	People in the risky area from SuratThani southward shoul***d beware of*** flooding conditions.
2.	in (freq.7,956)	cold in (178)	Northern Part Morning fog, temperature will change slightly; ***Cold in*** the upper part and high mountain, minimum temperature 14- 17 °c, elsewhere cool, minimum cool 18-20 °c.

Table 4.8 (Continued)

No.	Word	Grammatical Collocations	Examples
3.	to (freq.6,434)	scattered to (216)	***Scattered to*** almost widespread thundershowers with isolated heavy rain are expected over Thailand.
4.	with (freq.3,926)	cloudy with (369)	Northeastern Part from Jun 28 to Jul 1, ***cloudy with*** *scattered thundershowers* and isolated heavy rain mostly in the East Part.
5.	for (freq.3,107)	alert for (31)	This condition alerts for provinces of Kamphaengphet, Phetchabun, Trat, Ranong and Phangnga.
6.	at (freq.2,439)	water at (336)	Sun rises at 6.21 am Sun sets at 5.47 pm: Moon rises at 6.21 am Moon sets at 5.46 pm: High ***water at*** 10.24 am and at 5.52 pm: Low ***water at*** 3.15 am and at 2.17 pm
7.	from (freq.2,384)	calculated from (89)	Bangkok Air Quality Index: Worst Area Khlong Chan ***Calculated from*** dust smaller than 10 microns plus nitrogen dioxide, sulphur dioxide, ozone and carbon monoxide.

4.6 Compounds

In scientific language or technical discourse, it seems that multi-word vocabulary, in this case a complex noun phrase, plays an important role in carrying most contents (Martin, 2001). According to Chung and Nation (2003), technical vocabulary with the rating scale was based on the degree of technicalness of words in context. Chung and Nation found that many technical words are from common words, including words from the General Service Lists (GSL) and the

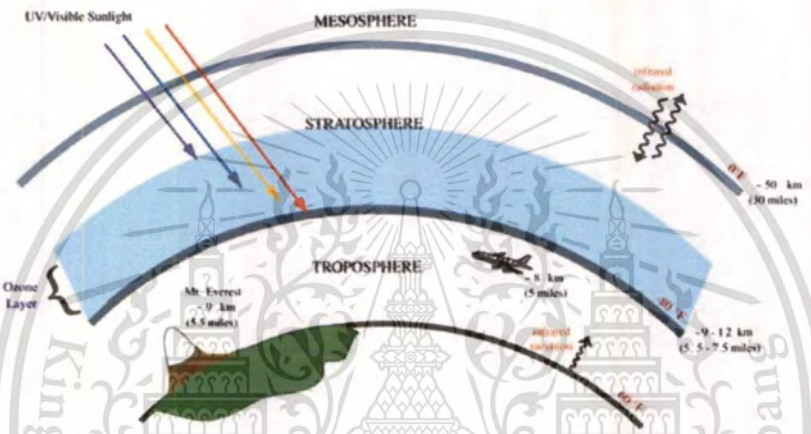
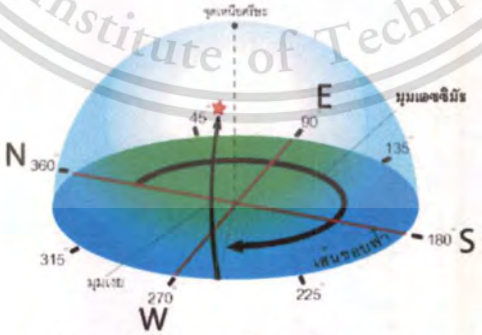
Academic Word Lists (AWL). Similarly, individual word in this study can be a part of a multi-word lexical unit as a complex noun phrase. For instance, the example of technical noun phrases (TNPs) containing the General Service Lists (GSL), the Academic Word Lists (AWL), and other vocabularies used in this study in the Weather Forecast Corpus as shown below.

- station temperature, extreme minimum temperature, sea surface temperature,
- heavy rain, widespread rain, fairly widespread rain, torrential rain, isolated rain,
- widespread thundershowers, isolated thundershowers, scattered thundershowers, widely scattered thundershowers, almost widespread showers,
- low pressure cell, high pressure system, heat wave, moderate sea,
- moderate southwest monsoon, northeasterly wind

Some technical noun phrases were created from the General Service Lists (GSL), the Academic Word Lists (AWL), and other vocabularies which are technically used with meteorological meanings. From these examples, it can be seen that words from the General Service Lists (GSL), the Academic Word Lists (AWL), and other vocabularies are not limited to providing meanings which are common to all academic texts. These words in the examples with other type of words outside technical words became technical such as extreme minimum temperature, widespread rain, fairly widespread rain, almost widespread showers, low pressure cell, torrential rain, isolated rain, northeasterly wind, widespread thundershowers, scattered thundershowers, etc. Some are used together with words in their groups and become technical, for examples, general vocabulary + general vocabulary (e.g. station temperature, sea surface temperature, heavy rain, high pressure system, heat wave, moderate sea) and other words + other words (e.g. isolated thundershowers). To conclude from the examples, many words from the General Service Lists (GSL), the Academic Word Lists (AWL), and other vocabularies can provide technical meaning when collocate with other words as technical compound.

4.7 The Sample Entries of the Meteorological Dictionary

From the evidence of the Weather Forecast Corpus (WFC), information about the frequencies of words is very important for vocabulary grading and selection. Therefore, the occurrence frequencies of the technical vocabularies, compounds, and abbreviations with high frequencies in this study are proposed to be contained in the sample of meteorological dictionary. There are forty-one headwords which occurred more than ninety-nine times in the WFC.

A
accumulative rainfall n. /เออะ-คิว-'มู-เล-ทิฟ 'เรน-ฟอล/ ปริมาณฝนสะสม ปริมาณฝนที่ได้จากการที่ฝนตกตลอดระยะเวลาหนึ่ง เช่น ปริมาณฝนตั้งแต่ วันนี้ 07.00 น. ถึง 07.00 น. ของวันพรุ่งนี้ **Ex.** Temporal variations in daily ***accumulative rainfall*** (mm) and the percentage of habitats in the stream areas were aqueous compared to the initial sampling period.
atmosphere n. /'แอท-โมซ-'เฟีย/ บรรยากาศ (adj.atmospheric)    อากาศที่อยู่รอบ ๆ ตัวเรา และที่หุ้มห่อโลกจากพื้นดินขึ้นไป จนถึงระดับที่สูงสุดในท้องฟ้า เป็นส่วนผสมของก๊าซต่าง ๆ รวมทั้งไอน้ำ ซึ่งระเหยมาจากแหล่งน้ำต่างๆด้วย **Ex.** The barometer cistern must be made airtight except for a lead to a special head exponsed to the ***atmosphere*** and designed to ensure that the pressure inside is true static pressure.
azimuth (Az) n. /'แอซ-ซิ-มัธ/ เส้นวางราบ    มุมตำแหน่ง หรือทิศทางของดวงดาวที่วัดบนพื้นโลก เป็นมุมในแนวราบ นับจากทิศเหนือ ในทิศทางตามเข็มนาฬิกา ไปยังทิศตะวันออก ทิศใต้ ทิศตะวันตก และกลับมาทิศเหนืออีกครั้งหนึ่ง ใช้ในการบอกทิศทาง **S.** bearing **Ex.** Electronic scanning in ***azimuth*** (over a limited sector) is accomplished by changing the frequency of beams coming out of a slotted waveguide.

B

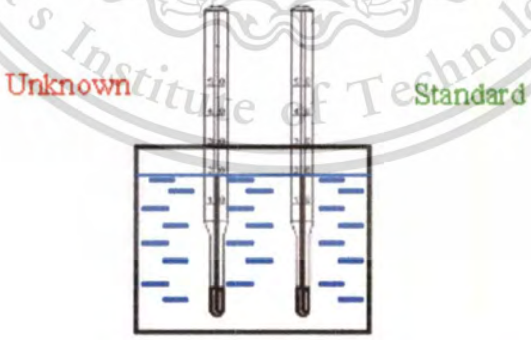
barometer n. /'บา-รอ-มิ-เทอะ/ เครื่อง (มาตร) วัดความกดอากาศ

เครื่องมือสำหรับวัดความกดของอากาศหรือบรรยากาศ มีสอง ชนิด คือ บารอมิเตอร์ปรอท (mercury barometer) หรือบารอมิเตอร์แบบคิว (kew barometer) และ บารอมิเตอร์ตลับ (aneroid barometer)

Ex. In order to maintain the acceptable performance of a *barometer*, the calibration correction applied to the reading must be checked at relatively frequent intervals.

C

calibration n. /'แค-ลิ-เบร-เฌิน/ การสอบเทียบมาตรฐานเครื่องมือที่ใช้ตรวจอากาศ



การปฏิบัติการเปรียบเทียบผลการวัดของเครื่องมือวัดอากาศที่ไม่รู้ค่าความถูกต้องกับเครื่องมือ หรือ อุปกรณ์มาตรฐานที่รู้ค่าความถูกต้องเพื่อหาความสัมพันธ์ระหว่างกัน และรายงานผลรวมทั้งดำเนินการปรับแต่งเครื่องมือวัดในกรณีที่ผลการวัดผิดไปจากเกณฑ์ที่กำหนด

Ex. During *calibration*, each of the sensing element can be checked with a resolution of 0.01 hPa.

Celsius (°C/°c) n. /'เซล-เซียซ/ องศาเซลเซียส

Kelvin temperature conversion formulae

	from Kelvin	to Kelvin
Celsius	[°C] = [K] − 273.15	[K] = [°C] + 273.15
Fahrenheit	[°F] = [K] × 9/5 − 459.67	[K] = ([°F] + 459.67) × 5/9
Rankine	[°R] = [K] × 9/5	[K] = [°R] × 5/9

For temperature *intervals* rather than specific temperatures,

1 K = 1 °C = 1.8 °F = 1.8 °R

สเกลหรือมาตราวัดอุณหภูมิเป็นหน่วยวัดอุณหภูมิหน่วยหนึ่งในระบบเอสไอ ใช้มากในด้านวิทยาศาสตร์ในทางอุตุนิยมวิทยานิยม

S. centigrade

Ex. Eastern Part Coastal Hot with maximum temperature 35-38 °c.

climate n. /'ไคล-เมท/ ภูมิอากาศ (adj. climatological)

ลักษณะเงื่อนไขของบรรยากาศ ที่อยู่บริเวณใกล้ผิวโลก ณ บริเวณใดบริเวณหนึ่ง ซึ่งได้จากการเฉลี่ยข้อมูลสภาพอากาศระยะยาว

Ex. NASA's Earth Science Enterprise is dedicated to applying Earth System Science to improve prediction of *climate*, weather, and natural hazard using the unique vantage point of space.

climate change n. /'ไคล-เมท-เชนจ/ การเปลี่ยนแปลงสภาพภูมิอากาศ

ความแปรปรวนอย่างมีนัยยะสำคัญ ทางสถิติของสภาวะภูมิอากาศ หรือ ความแปรปรวนที่เกิดขึ้นอย่างถาวรเป็นเวลานาน (ทศวรรษ หรือนานกว่า นั้น) การเปลี่ยนแปลงภูมิอากาศ เกิดขึ้นเนื่องจากกระบวนการภายในหรือแรงกดดันภายนอก หรือเกิดจาก การเลี่ยนแปลงของบรรยากาศ หรือ พื้นดิน ที่เกิดจากกิจกรรมต่อเนื่องของ มนุษย์

S. climatic change

Ex. Linking *climate change* with broader sustainable consumption and production concerns, human rights issues, and democratic values is crucial for shifting societies toward more sustainable development pathways.

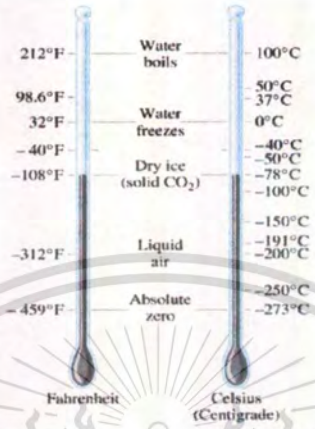
cloudy adj. /'คลาว-ดี/ ลักษณะที่ปกคลุมด้วยเมฆ

ลักษณะที่เต็มไปด้วยก้อนเมฆที่ปกคลุมท้องฟ้า

Ex. Chiang Mai - *Cloudy* with almost widespread thundershowers and isolated heavy rain.

F

Fahrenheit (°F) n. /'ฟา-เรน-ไฮท/ องศาฟาเรนไฮต์



สเกลหรือมาตราวัดอุณหภูมิเป็นหน่วยวัดอุณหภูมิหน่วยหนึ่ง มีจุดเยือกแข็งอยู่ที่ 32 องศาฟาเรนไฮต์ โดยปกติจะเขียนว่า 32 °F และมีจุดเดือดที่ 212 องศาฟาเรนไฮต์

Ex. Minimum temperature 23 °C (73°*F*). Maximum temperature 32 °C (90 °*F*).

fog n. /ฟือก/ หมอก (adj. foggy)

น้ำในอากาศหรือไฮโดรมีทีเออร์ (hydrometeor) ชนิดหนึ่งที่ประกอบด้วยกลุ่มละอองน้ำขนาดเล็กมาก สามารถมองเห็นได้ด้วยตาเปล่าลอยอยู่ในอากาศใกล้พื้นดิน โดยปกติจะทำให้ทัศนวิสัยทางแนวนอนที่ผิวพื้นโลกลดลง

Ex. Northern Part - Cool with *fog* in the morning throughout of the period.

forecast n. / 'ฟอร-คาซท / การพยากรณ์อากาศ v. พยากรณ์อากาศ

การคาดหมายลักษณะของบรรยากาศในอนาคตจากลักษณะของสารประกอบอุตุนิยมวิทยา เช่น หยาดน้ำฟ้า เมฆ ลม และอุณหภูมิ โดยนักอุตุนิยมวิทยา

S. (n.) prediction, (v.) predict

Ex. Today: *Forecast* for Maximum Temperature 32 °c: Sun sets at 6.34 p.m: High water at 5.04 a.m. and 7.30 a.m.

G

gust n. /กัซท/ ลมกระโชก (adj. gusty)

ลมแรงที่เกิดขึ้นทันทีทันใดชั่วขณะผิดไปจากความเร็วลมเฉลี่ยในเวลานั้น ระยะเวลาที่เกิดลมชนิดนี้โดยปกติมักจะต่ำกว่า 20 วินาที เมื่อลมชนิดนี้หมดไปแล้วความเร็วลมจะลดลงกลายเป็นลมอ่อน อาจหมายถึงความเร็วลมที่มีทิศแน่ชัดแต่กำลังลมอาจจะเพิ่มขึ้น หรือลดลงจากค่าปานกลางในช่วงเวลาจำกัด

Ex. Eastern Thailand - Scattered thundershowers with isolated heavy rain and *gust* mostly in Prachinburi, Sakaeo, Chachoengsao, Chonburi, Chanthaburi and Trat.

H

haze n. /เฮซ/ (adj. hazy) เมฆหมอก

อนุภาคเล็กๆที่ลอยอยู่ในอากาศ หรือการกระจัดกระจายของแสงที่ลดความสามารถในการมองเห็น หรือการผสมกันของละอองของเหลวและหมอกปนควันแบบโฟโตเคมี

Ex. Northern Thailand – Morning cold in the upper part while hot and *haze* during the day.

heat wave n. /'ฮีท เวฟ/ คลื่นอากาศร้อน

ช่วงระยะเวลาหนึ่งที่อากาศร้อนกว่าปกติ ซึ่งอาจมีความชื้นมาก และ เป็นการที่อุณหภูมิอากาศลดลงโดยสัมพันธ์กับการแผ่รังสีในเวลากลางคืน

Ex. More than 500 deaths in Hungary were attributed to the *heat wave*, while major and widespread wildfires destroyed large areas of forest across the region.

hectopascal (HPA/hPa) n. /'เฮค-โต-ปาส-กาล/ หน่วยของความกดอากาศ

หน่วยระหว่างประเทศเพื่อการวัดความกดอากาศหรือความดันบรรยากาศ (1 เฮคโตปาสกาล (hPa) = 100 ปาสกาล (Pascals) = 1 มิลลิบาร์ (mb))

S. millibar (mb)

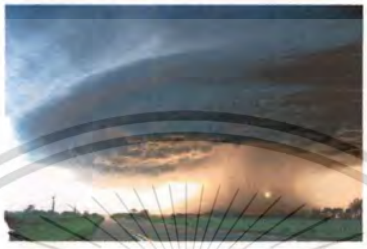
Ex. Any electronic barometer exceeds 0.25 *hPa* should be regarded as unserviceable.

humidity n. /'ฮิว-มิ-ดิ-ตี้/ ความชื้น (adj. humid)

สภาวะที่อากาศมีไอน้ำปะปนอยู่ ปริมาณไอน้ำในอากาศจะมากหรือน้อยขึ้นอยู่กับอุณหภูมิ ถ้าอุณหภูมิสูง อากาศจะรับไอน้ำได้มากกว่าอากาศที่มีอุณหภูมิต่ำกว่า

Ex. Iron a massive 50-year government database of ground observations and measurements collected by weather balloons, they obtained the temperature, pressure, ***humidity***, and winds at different height in the atmosphere.

hurricane n. /'เฮอ-ริ-เคน/ พายุหมุนเขตร้อนเฮอริเคน



คำเรียก พายุหมุนเขตร้อน ที่เกิดแถบมหาสมุทรแอตแลติก มีความเร็วลมสูงสุดใกล้ศูนย์กลาง 65 นอต หรือมากกว่า หรือตั้งแต่ 118 กิโลเมตรต่อชั่วโมงขึ้นไป

Ex. Texas Governor allowed Texas to launch preparations such as pre-deploying resources to ensure local communities are ready to respond to the ***hurricane***.

K

kelvin (K) n. /'เคล-วิน/ หน่วยอุณหภูมิสัมบูรณ์

Kelvin temperature conversion formulae

	from Kelvin	to Kelvin
Celsius	[°C] = [K] − 273.15	[K] = [°C] + 273.15
Fahrenheit	[°F] = [K] × 9/5 − 459.67	[K] = ([°F] + 459.67) × 5/9
Rankine	[°R] = [K] × 9/5	[K] = [°R] × 5/9

For temperature *intervals* rather than specific temperatures,

1 K = 1 °C = 1.8 °F = 1.8 °R

สเกลหรือมาตราวัดอุณหภูมิเป็นหน่วยวัดอุณหภูมิหน่วยหนึ่งในระบบเอสไอ

Ex. For a temperature of 300 K, 99.99 percent of the power of the terrestrial radiation has a wave length longer than 300 nm and about 99 percent longer than 5000 nm.

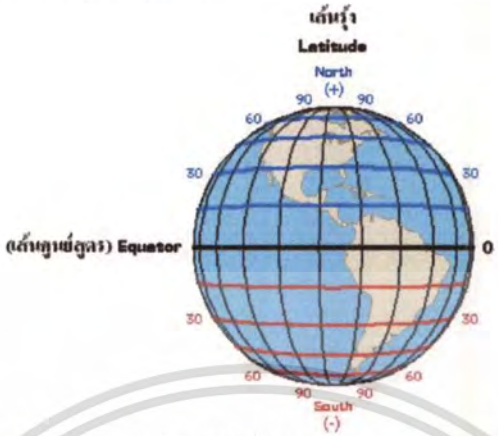
knot (kt/KT) n. /น็อต/ หน่วยวัดความเร็วลม

หน่วยความเร็วลมในระบบการเดินเรือ 1 ไมล์ทะเลจะเท่ากับ 1.1508 หรือ 1.852 กิโลเมตรต่อชั่วโมง หรือ 1.687 ฟุต (0.5144 เมตรต่อวินาที)

Ex. The Gulf of Thailand (Area1-and 2) – Scattered thundershowers. Southwesterly winds 11-18 ***knots*** or 20-35 km/hr.

L

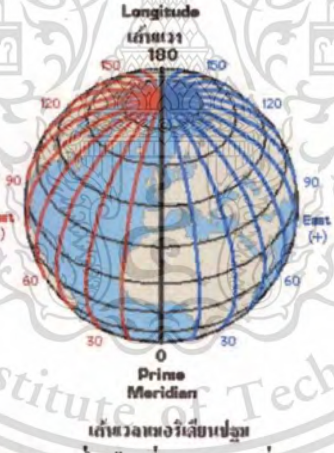
latitude (Lat/lat) n. /'แลต-ติ-ทยูด/ เส้นรุ้ง/เส้นขนาน



ระยะทางเชิงมุมที่วัดไปทางเหนือและใต้ของเส้นศูนย์สูตร นับจาก 0 องศา ไปทางเหนือและทางใต้ 90 องศา เส้นรุ้งหรือเส้นขนานมีเส้นศูนย์สูตรเป็นเส้นอ้างอิงตามธรรมชาติ

Ex. The tropical storm "Noul" has made landfall at Nha-Trang, Vietnam was centered about *lat* 11.5 °N, long 108.8 °E, with sustained winds of 75 km/hr or 40 knots.

longitude (Long/long) n. /'ลื่อง-จี้-ทยูด/ เส้นแวง



ระยะทางเชิงมุมที่วัดจากเส้นเวลาเมริเดียนปฐมซึ่งถือที่ 0 องศา ที่ตำบลกรีนิชเป็นหลัก วัดไปทางตะวันออก 180 องศาตะวันออก และทางตะวันตก 180 องศาตะวันตก เส้นเวลากรีนนิช เป็นเส้นเวลาเมอริเดียนปฐมเป็นเส้นสมมติขึ้นมาหนึ่งเส้นสำหรับอ้างอิง และเป็นเส้นแวง 0 องศา

Ex. The tropical storm "Lekima" was centered about 400 km, southeast of Danang, Vietnam or Lat 16.3 °N, *Long* 111.8 °E, with maximum sustained winds of 55 knots or 100 km/hr.

Lunar Standard Time (LST) n. /'ลู-นะ 'ซแทน-เดิรด ไทม/ เวลามาตรฐานตามจันทรคติ เวลามาตรฐานตามเงื่อนไขของดวงจันทร์ซึ่งนำมาเปรียบเทียบในเวลาบนโลก

Ex. Daily Weather Forecast Distribution on Thursday 13 September 2007 Time issued 1700 *LST*

M

mercury barometer n. /'เมอ-คิว-รี 'บา-รอ-มิ-เทอะ/ บารอมิเตอร์ปรอท

บารอมิเตอร์ซึ่งใช้หลักขั้นมูลฐาน คือ ความกดอากาศหรือบรรยากาศจะต้องสมดุลย์กับน้ำหนักของปรอทในลำหลอดแก้วที่คว่ำอยู่ในอ่างปรอท ดังนั้น ความสูงของปรอทในลำหลอดแก้วจะขึ้นอยู่กับความกดอากาศ บารอมิเตอร์อาจแบ่งออกได้เป็น 3 พวก ตามหลักการสร้างของแต่ละชนิด คือ พวกที่สร้างเป็นแบบกระปุกปรอท (cistern barometers) แบบกาลักน้ำ (siphon barometers) และแบบชั่งน้ำหนัก (weight barometers)

S. mercurial barometer, kew barometer

Ex. If a station has one *mercury barometer* and one electronic barometer, it would be normal for the mercury barometer to be the low-drift barometer.

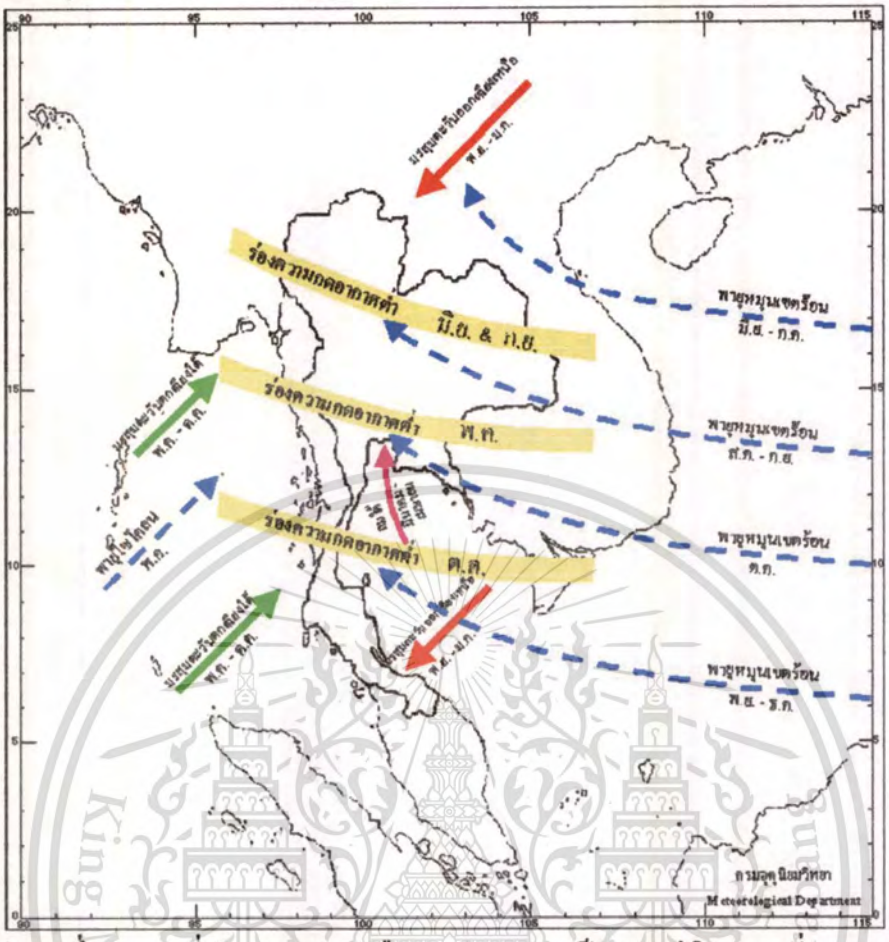
meteorology n. /มี-ทิ-ออ-'รอล-เลาะ-จี้/ อุตุนิยมวิทยา (adj. meteorological)

วิชาที่กล่าวถึงเรื่องราวของบรรยากาศ รวมทั้งความสัมพันธ์ระหว่างบรรยากาศกับพื้นโลก มหาสมุทร และสิ่งมีชีวิตโดยทั่วไป วิชานี้เกี่ยวข้องกับลักษณะทางกายภาพ ทางเคมี และการเปลี่ยนแปลงของบรรยากาศ จุดมุ่งหมายของวิชานี้อยู่ที่การศึกษาให้เข้าใจในเรื่องราวของบรรยากาศ การพยากรณ์อากาศให้เป็นไปโดยสมบูรณ์ถูกต้องและแม่นยำ

S. atmospheric science

Ex. The observation of clouds and the estimation of the height of their bases above the Earth's surface are important aviation and other operational applications of *meteorology*.

monsoon n. /'มึ่อน-ซูน/ ลมมรสุมหรือมรสุม (adj. monsoonal)

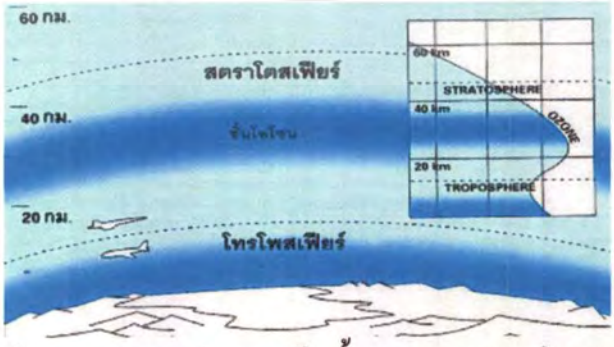


การหมุนเวียนส่วนหนึ่งของลมที่พัดตามฤดูกาล คือ ลมประจำฤดู เป็นลมแน่นิ่งและสม่ำเสมอ เกิดขึ้นเนื่องจากความแตกต่างระหว่างอุณหภูมิของพื้นดินและพื้นน้ำทำนองเดียวกับลมบกลมทะเล

Ex. Abundant rains and isolated heavy rain are likely in the windward side of the *monsoon* or western portion including the East of the country.

O

ozone (O3) n. /'โอ-โซน/ โอโซน



โมเลกุลที่ประกอบจากออกซิเจน 3 อะตอม ปรากฏอยู่ในชั้นบรรยากาศของโลก และมีการใช้งานในทางอุตสาหกรรมและเครื่องใช้ตามบ้านทั่วไป

Ex. Fahey received an award for his work on many aspects of stratospheric ozone depletion and the impact of aviation on *ozone* and climate.

P

precipitation (Precip/precip) n. /พริ-ซิพ-'พิ-เท-เฌิน/ น้ำฟ้าหรือหยาดน้ำฟ้า

น้ำที่มีลักษณะเป็นของเหลว หรือของแข็งรูปผลึก ซึ่งเกิดจากก้อนเมฆบนท้องฟ้าแล้วตกลงมายังพื้นโลก เช่น ฝน หิมะ ลูกเห็บ ลักษณะของหยาดน้ำฟ้าดังกล่าวแล้วแตกต่างไปจากเมฆ หมอก น้ำค้าง น้ำค้างแข็ง และไอน้ำ หรือน้ำแข็งในรูปอื่นๆ ตรงที่หยาดน้ำฟ้าจะต้องตกจากบรรยากาศถึงพื้นดิน

Ex. Systematic losses will vary by type of *precipitation* (snow, mixed snow and rain, rain).

R

radiation n. /เร-'ดี้-เอ-เฌิน/ การแผ่รังสี

1.กระบวนการที่รังสีแม่เหล็กไฟฟ้าแพร่กระจาย ผ่านช่องว่างการที่พลังงานถูกส่ง (emission) หรือถ่ายเท (transfer) ออกไปในรูปของแม่เหล็กไฟฟ้า หรือในรูปของการส่งอนุภาค

2. ตัวพลังงานเองเรียกว่า "พลังงานจากการแผ่รังสี" (radiant energy)

Ex. Light is the *radiation* visible to the human eye.

rainfall n. /'เรน-ฟอล/ ปริมาณน้ำฝน

ปริมาณหยาดน้ำฟ้าที่เป็นของเหลว (ฝน) ปกติจะวัดได้โดยถังวัดน้ำฝน

Ex. The low pressure trough lies across the upper northern and northeastern, more *rainfall* is likely in the areas.

relative humidity (RH) n. /'เระ-เลอะ-ทิฟ 'ฮยู-มิ-ดิ-ตี้/ ความชื้นสัมพัทธ์

Daily Observation Report

Friday, April 15, 2011 at 7AM

Issued On 4/15/2011 7:31:40 AM

Station	PPP (hPa)	T (c)	Tmax (c)	dTmax (24h)	Tmin (c)	dTmin (24h)	R24hr (mm)	R1Jan (mm)	RH (%)	Wind (deg)	Wind (knot)
Northern Part											
MAE HONG SON	1009.29	24.0	36.2	2.0	23.5	2.1	0.0	94.0	86	C	0
MAE SARIANG	1010.00	23.3	34.3	0.7	22.0	0.9	0.0	174.7	94	C	0
CHIANG RAI	1009.50	22.3	31.9	6.4	21.7	2.8	0.0	150.2	95	C	0

อัตราส่วนระหว่างมวลของไอน้ำที่มีอยู่จริงในอากาศขณะนั้นกับมวลของไอน้ำอิ่มตัวที่อุณหภูมิและปริมาตรเดียวกัน มีหน่วยเป็นร้อยละ

Ex. Now in Bangkok Temperature 32.0 °c Dewpoint 19.0°c *Relative humidity* 46%

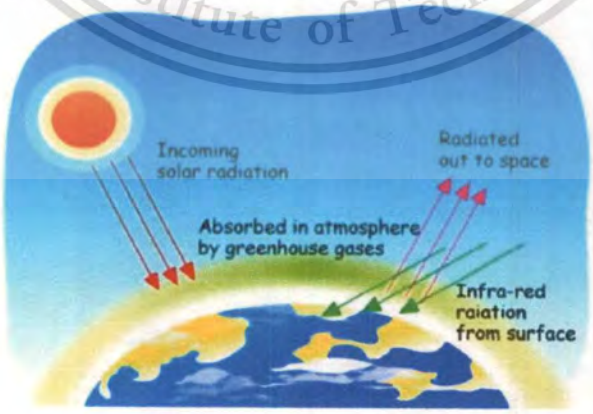
S

sea ice n. / 'ซี ไอซ/ ทะเลน้ำแข็ง

ลักษณะของน้ำแข็งที่เกิดจากการแข็งตัวของน้ำทะเล

Ex. The variables sent from the *sea ice* back to the atmosphere are sea surface temperature, sea ice fraction, and thickness and snow depth.

solar radiation n. /'โซ-เลอะ เร-'ดี้-เอ-เฌิน/ การแผ่รังสีของดวงอาทิตย์



รังสีแม่เหล็กไฟฟ้าที่ปล่อยออกจากดวงอาทิตย์

Ex. Smoke in the atmosphere temporarily reduces the amount of *solar radiation* reaching the surface.

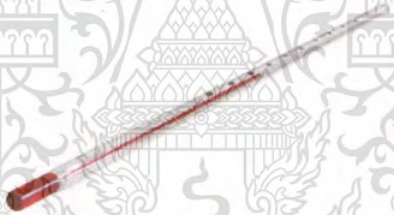
southwest monsoon n. /'ซาวธ-เวซท 'มึ่อน-ซูน/ มรสุมตะวันตกเฉียงใต้

มรสุมที่พัดปกคลุมประเทศไทยระหว่างกลางเดือนพฤษภาคมถึงกลางเดือนตุลาคม โดยมีแหล่งกำเนิดจากบริเวณความกดอากาศสูงในซีกโลกใต้ บริเวณมหาสมุทรอินเดีย ซึ่งพัดออกจากศูนย์กลางเป็นลมตะวันออกเฉียงใต้และเปลี่ยนเป็นลมตะวันตกเฉียงใต้เมื่อพัดข้ามเส้นศูนย์สูตร ซึ่งเป็นมรสุมที่จะนำมวลอากาศชื้นจากมหาสมุทรอินเดียมาสู่ประเทศไทยทำให้มีเมฆมากและฝนตกชุกทั่วไป

Ex. Strong *southwest monsoon* prevails over the Andaman Sea, Thailand and the Gulf.

T

thermometer n. /'เธอ-มอ-มิ-เทอะ/ เครื่องมือวัดอุณหภูมิ



เครื่องมือที่ใช้ในการตรวจวัดอุณหภูมิ หรือระดับความร้อน ความเย็น เทอร์โมมิเตอร์แบ่งออกได้เป็น หลายแบบ แล้วแต่การสร้างหรือชนิดที่ใช้

Ex. Sufficient measurements should be taken to ensure that the corrections to be applied represent the performance of the *thermometer* under normal conditions.

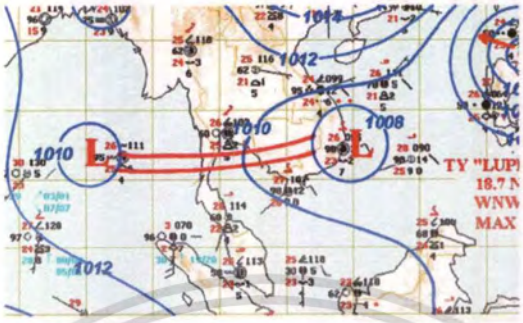
thunderstorm n. /'ธัน-เดอ-ซตอม/ พายุฟ้าคะนอง

พายุที่เกิดเฉพาะท้องถิ่น เกิดจากเมฆคิวมูโลนิมบัส มีฟ้าแลบกับฟ้าร้องรวมอยู่ด้วย นอกจากนั้นมักจะมีลมกระโชกแรงและฝนตกหนักเกิดขึ้น บางครั้งยังมีลูกเห็บตกลงมาด้วย เป็นพายุที่เกิดขึ้นในช่วงเวลาอันสั้น มีน้อยครั้งที่เกิดขึ้นนานกว่า 2 ชั่วโมง

S. electric storm

Ex. The moderate high pressure from China has extended to cover northeastern and northern Thailand, cause *thunderstorm*, gusty and decreasing in temperature.

trough (L) n. /ทร็อฟ/ ร่องความกดอากาศต่ำ



แถบหรือแนวของ อากาศซึ่งมีความกดต่ำเมื่อเทียบกับบริเวณรอบๆในระดับเดียวกัน ปรากฏบนแผนที่อากาศในรูปแบบของระบบเส้นความกดเท่าหรือเส้นไอโซบาร์ หรือเส้นความสูงเท่า เกือบเป็นรูปตัววี (V) โค้งแหลมยื่นออกไปจากดีเปรสชันหรือบริเวณความกดอากาศต่ำ เส้นเหล่านี้เกือบจะขนานกัน

Ex. A *trough* of low pressure lies across lower northern, central and northeastern Thailand to a low pressure cell in South China Sea.

U

Universal Time Coordinated (UTC) n. /ยู-นิ-'เวอ-เซิล โค-'ออร-ดิ-เนท ไทม/ เวลาสากลเชิงพิกัด

หน่วยเวลาที่ใช้ในการอ้างอิงการหมุนของโลก โดยใช้เครื่องหมาย บวก (+) หรือ ลบ (-) เทียบจากหน่วย เวลาสากล ซึ่งเป็นระบบอ้างอิงจาก เวลามาตรฐานกรีนิช (UTC/GMT) จุดอ้างอิงของเวลาสากลเชิงพิกัดคือที่เส้นแวง ที่ 0° ที่ตัดผ่าน Royal Greenwich Observatory ใน กรีนิช, ลอนดอน, สหราชอาณาจักร (และเป็นสาเหตุหลักที่เวลา UTC ยังคงมีใช้อยู่ในปัจจุบัน)

S. Greenwich Mean Time (GMT)

Ex. BROADCAST MESSAGE FOR SHIPPING Synoptic situation at 13-0900 *UTC*. Rather strong northeast monsoon prevails over the Gulf of Thailand.

V

visibility n. /วิซ-ซะ-'บิล-ลิ-ที/ ทัศนวิสัย

ระยะทางไกลที่สุดในทิศทางที่กำหนดไว้ ซึ่งผู้ตรวจสามารถมองเห็นวัตถุที่มีขนาดพอสมควรได้ด้วยตาเปล่า และบอกได้ว่าวัตถุนั้นเป็นอะไร โดยวัดค่าทัศนวิสัยออกมาเป็นระยะทาง มีหน่วยเป็นเมตร กิโลเมตร และไมล์

Ex. Morning fog remains in all areas and all forms of transportation should beware of poor *visibility*.

W

weather forecast bureau n. /'เวธ-เธอะ 'ฟอร-คาซท 'บิว-โร/ สำนักพยากรณ์อากาศ

สำนักงานหรือหน่วยงานบริหารงานด้านวิทยาศาสตร์แห่งชาติหรือส่วนภูมิภาคของกรมอุตุนิยมวิทยาที่มีกิจกรรมเกี่ยวข้องกับด้านอุตุนิยมวิทยาในการพยากรณ์อากาศทั้งภาคทฤษฎีและภาคปฏิบัติ

Ex. All ship should proceed with caution and small boat in the Andaman Sea should keep ashore.

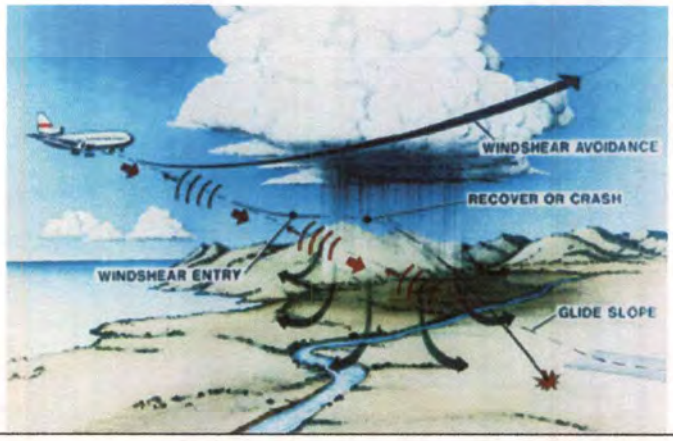
Weather Forecast Bureau, Meteorological Department 14 July 2006 04.00 p.m.

weather outlook n. /'เวธ-เธอะ 'อาวท-ลุค/ รายงานสถานการณ์ของสภาวะอากาศ

รายงานสภาพอากาศที่ระบุแนวโน้มภัยธรรมชาติหรือสถานการณ์น้ำ เพื่อที่จะให้ข้อมูลข่าวสารสำหรับผู้ที่เกี่ยวข้องได้เตรียมตัวป้องกันภัย

Ex. *Weather outlook* for the week on July 10-16, 2006, rather strong southwest monsoon prevails over the Andaman Sea, Thailand and the Gulf.

wind shear n. /'วินด-เฉีย/ ลมเฉือน



กระแสลมที่เกิดการเปลี่ยนแปลงของความเร็วและทิศทางในระยะสั้นๆ แต่รวดเร็วเฉียบพลันซึ่งเกิดขึ้นได้ทั้งแนวราบและแนวตั้งในทุกระดับความสูง ตั้งแต่พื้นดินถึงชั้นบรรยากาศ

S. shear vector

Ex. Increased *wind shear* in an El Niño tends to rip apart embryonic storms in the Atlantic basin and Caribbean Sea.

World Meteorological Organization (WMO) n. /เวิลด มี-ทิ-ออ-'รอล-เลาะ-จิ-คัล ออ-กา-ไน-'เซ-เฌิน/ องค์การอุตุนิยมวิทยาโลก

องค์การชำนาญพิเศษขององค์การสหประชาชาติ มีวัตถุประสงค์เพื่ออำนวยความสะดวกในการร่วมมือประสานงานกัน ส่งเสริมการวางมาตรฐานในการทำงานและปรับปรุงกิจกรรมอุตุนิยมวิทยาทั่วโลก และเพื่อสนับสนุดสมรรถภาพในการแลกเปลี่ยนข่าวสารระหว่างประเทศในด้านผลประโยชน์ต่อกิจกรรมต่างๆของมวลมนุษย์

Ex. Over the past four years, the *WMO* President came from Russia.

4.8 Dicussions on the Results of this Study

4.8.1 The Vocabulary in the Weather Forecast Texts

Regarding types of vocabulary in this study, the results showed that the vocabulary in the weather forecast texts was mainly general. The highest major open class words used in the Weather Forecast Corpus (WFC) is noun. The General Vocabulary (GV) has the highest proportion (76.37 % of tokens) in this study because the weather forecasts were broadcasted to the general public. Therefore, the general vocabularies were often used in the weather forecasts in

this study. Some general vocabularies were found to be used with meteorological meaning such as 'Hot' (scale of temperature from 35.0 to 39.9 °C), 'Cool' (scale of temperature from 18.0 to 22.9 °C), 'Cold' (scale of temperature from 8.0 to 15.9 °C), 'Scattered' (the amount rainfall from 40 % but not more than 60 % of a forecasted-area or a sky coverage of 1/8 through 4/8. In U.S. weather observing procedures, this is reported with the contraction 'SCT'), 'Widespread' (the amount rainfall from 80 % up of a forecasted-area), 'Calm' (wave height from 0.0 m. to 0.10 m.), and 'Smooth' (wave height from 0.10 m. to 0.50 m.). In terms of the Other Vocabulary, it had ahigh proportion (12.58 % of tokens) in the Weather Forecast Corpus. The most frequent vocabularies in the group of the Other Vocabulary contained many proper nouns concerned with forecasted-areas of Thailand weather forecasts, for example, '*Northern Thailand-Cool with morning fog in the upper portion. Scattered thundershowers and isolated heavy rain in* ***Phrae, Nan, Uttaradit, Sukhothai*** *and* ***Phitsanulok.***

To sum up, this study indicated that language of weather forecast text is not highly technical compared to other technical corpora such as Pharmaceutical Leaflets Corpus (Pongnan, 2006). This finding does not support the view of Chung and Nation (2003) that technical words should be much more frequent in the technical corpus.

4.8.2 The Collocations in the Weather Forecast Corpus

Language for Meteorology contains specialized collocations. There are seven forms of the lexical collocations found in the WFC: 1) Adjective + Noun, 2) Noun + Noun, 3) Noun + Noun + Noun, 4) Adjective + Noun + Noun, 5) Adjective + Adjective + Noun, 6) Adverb + Noun, and 7) Adverb + Adjective + Noun. The form of 'Adjective + Noun' often occur in the WFC. Most of them concern with types of rain and temperature such as 'scattered thundershowers', 'widespread thundershowers, and 'isolated thundershowers.' The word 'scattered' is an adjective which can be used with the words 'scattered thundershowers', 'scattered thunderstorms', 'scattered rain', and 'scattered showers' or can be used the adjectives 'isolated' and 'widespread' with the types of rain instead of the 'scattered'. Table 4.9 shows the forms of the lexical collocations in the Weather Forecast Corpus as follows:

Table 4.9 Forms of Lexical Collocations in the Weather Forecast Corpus

No.	Word Classes Forms of Lexical Collocations	Examples
1.	Adjective + Noun	scattered thundershowers (freq.2,048)
		maximum temperature (freq.1,282)
		minimum temperature (freq.1,264)
		widespread thundershowers (freq.482)
		isolated thundershowers (freq.468)
		isolated showers (freq.89)
		gusty wind (freq.62)
		high wave (freq.61)
		scattered thunderstorms (freq.34)
		scattered rain (freq.30)
		scattered showers (freq.27)
		geopotential height (freq.9)
2.	Noun + Noun	wave height (freq.2,016)
		weather forecast (freq.147)
		mean temperature (freq.72)
		average temperature (freq. 53)
		wind shear (freq.30)
		solar wind (freq.26)
		heat wave (freq.16)
		cloud height (freq.13)
		ceiling height (freq.8)
3.	Noun + Noun + Noun	weather forecast bureau (freq.111)
		strong wind-wave (freq.17)
4.	Adjective + Noun+ Noun	daily weather forecast (freq.371)
		isolated heavy rain (freq.683)
		widely scattered thundershowers (freq.553)

Table 4.9 (Continued)

No.	Word Classes Forms of Lexical Collocations	Examples
5.	Adjective + Adjective + Noun	almost widespread thundershowers (freq.377)
		isolated heavy fall (freq.115)
		isolated light rain (freq.40)
6.	Adverb + Noun	northeasterly wind (freq.683)
		easterly wind (freq.225)
		southerly wind (freq.57)
7.	Adverb + Adjective + Noun	widely scattered thundershowers (freq.553)
		almost widespread thundershowers (freq.377)

4.8.3 Compounds and Multiword Terms in the Weather Forecast Corpus

The Weather Forecast Corpus shows that most of technical terms are compounds and multiword terms. The compounds were created from general words and consisted completely of nouns, with preceding nouns functioning as adjectives (Durham, 1994: 76). They are technically used with meteorological meanings such as '*station temperature*', '*soil temperature*', '*sea surface temperature*'.

4.8.4 Abbreviations and Symbols in Weather Forecasts

In Weather Forecast Corpus, there were six main kinds of abbreviations, i.e. clipping, initialism, contraction, substitution, acronym, and symbol. They took 10.28 % or 30,784 tokens of text coverage in this study. The results indicated the importance of the abbreviations or symbols in the daily weather forecasts. Most of the abbreviations refer to unit of measurement of temperature such as '°C' (Celsius) and '°F' (Fahrenheit), types of temperature: max (Maximum temperature), min (Minimum temperature), time (e.g. hr. (Hour), P.M./p.m. (Post Meridiem), A.M./a.m. (Ante Meridiem), UTC (Universal Time Coordinated), min (Minute)), wind (e.g. hr. (Hour) and height of wave (e.g. m (Meter/Metre), H (Height)). Examples of the daily weather forecasts from the TMD and the BP are as follows:

Weather Forecast For Thailand From 6:00 AM Wednesday To 6:00 AM Thursday.	
Northern Part	Scattered thundershowers and isolated heavy rain mainly in Chiang Mai, Lampang, Lamphun, Sukhothai, Tak, Phitsanulok, Kamphaeng Phet and Phetchabun. Minimum temperature 23-24 °C. Maximum temperature 31-34 °C. Easterly winds 10-30 km/hr.
Northeastern Part	Scattered thundershowers and isolated heavy rain mainly in Nakhon Ratchasima, Buri Ram, Surin, Si Saket and Ubon Ratchathani. Minimum temperature 22-25 °C. Maximum temperature 31-33 °C. Easterly winds 15-30 km/hr.
Central Part	Fairly widespread thundershowers and isolated heavy rain mainly in Nakhon Sawan, Lop Buri, Saraburi, Suphan Buri and Phra Nakhon Si Ayutthaya. Minimum temperature 23-24 °C. Maximum temperature 32-34 °C. Variable winds 15-30 km/hr.
Eastern Part	Fairly widespread thundershowers and isolated heavy rain mainly in Nakhon Nayok, Prachin Buri, Chanthaburi and Trat.

Figure 4.12 Daily Weather Forecast of Thailand in Thai Meteorological Department (TMD)

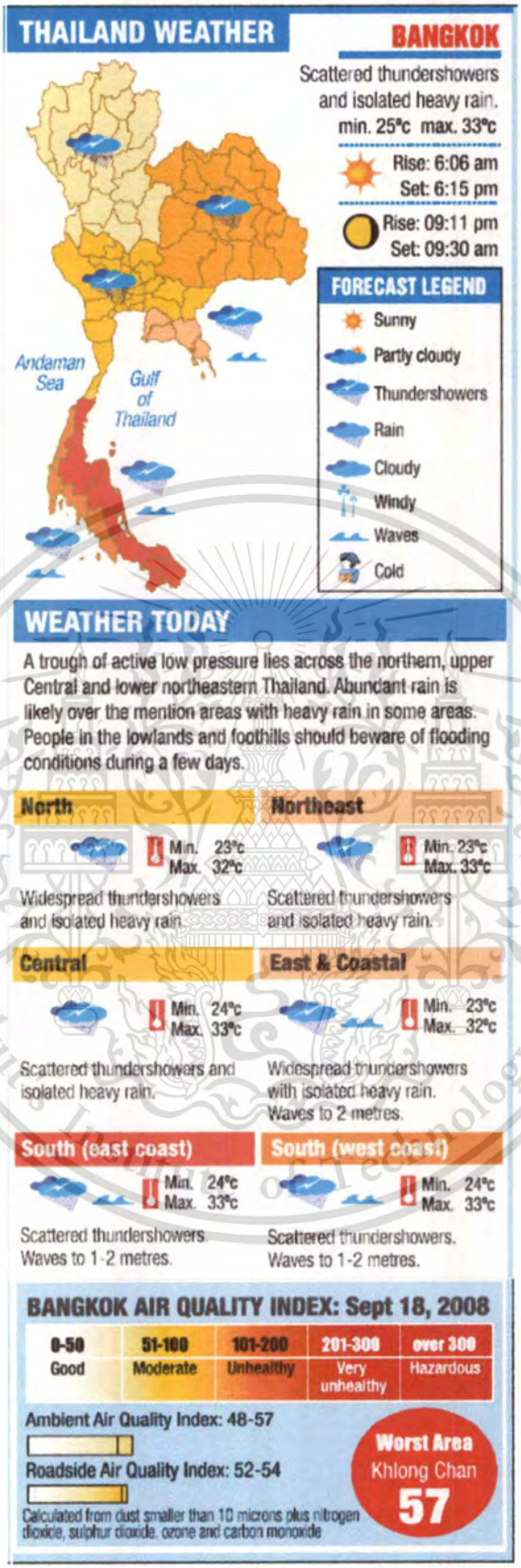


Figure 4.13 Daily Weather Forecast of Thailand in the Bangkok Post (BP) Newspaper

4.8.5 Limitations of the Corpus Written by Non-native English Speaker

There are some areas where the use of corpus can cause problems. The majority of corpora are based on authentic language use and this can be challenging for non-native speakers of English because most technical documents conform to fairly standard document types that present information in standard formats (Lebauer, 1985). Good technical communication should be accurate, clear, concise, coherent, and appropriate. The Weather Forecast Corpus was collected from non-native speakers of English (Thais). Therefore, there were some mistakes, i.e. ***'During January 6-8'.*** For this reason, we should be careful about specialized texts and the examples of sentences in the corpora which were not written by native speakers because they might contain problems with English grammar and technical phrase in sentences.

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS

This chapter is divided into three main parts. The first part describes the conclusions of this study. The second part presents the implications of the study. The final part focuses on the suggestions for further studies.

5.1 Conclusions of the Study

This study aimed to analyze vocabulary used in English weather forecast texts. The samples of this study were the English weather forecast texts which were divided into 2 main types: 1) weather forecasts (e.g. warning advisories, daily forecast, shipping forecast, weekly forecast, monthly forecast, sunrise-sunset and moonrise-moonset, and water level) and 2) meteorological documents (e.g. weather news , weather reports, meteorological textbook and manuals, meteorological journals, meteorological magazines, and other meteorological documents) during the year 2003-2009 in order to compile a Weather Forecast Corpus (WFC) which contains 555,818 words (tokens). The 13 sub-types of the texts were scanned and stored as plain text file (*.txt) in the corpus by using Optical Character Recognition (OCR). They were sent out into Microsoft Word 2003 as a word document file and were spell checked by the spell-checker. The research instrument of the study was the concordance software "Wordsmith Tools Version 3.0" (Scott, 1998). The statistical results in this corpus study were frequencies and percentages. All words in this study were divided into three groups. The first group, general vocabularies were classified with the General Service List (GSL) created by Bauman and Culligan (1995). The second group, academic vocabularies were classified with the Academic Word List (AWL) gathered by Coxhead (2000). The third group, technical vocabularies were identified by meteorological glossaries and the word outside those three groups. The guidelines on the technical vocabularies analysis were based on 'the Online Glossary of Meteorological Terms' (English-Thai) written by Thai Meteorological Department (2006), 'the Glossary of Meteorological Terms written by Thai Meteorological Department (1979), and 'the Online *Glossary of Meteorology*' written by American Meteorological Society (2000). The guidelines on

the abbreviations analysis were based on 'the Abbreviations Dictionary' written by Ralph De Sola et al. (1995).

The results showed all word lists (groups of vocabulary) of the WFC in the form of word frequency and percentage. According to statistical analysis, the whole corpus consists of 555,818 tokens. The total word types of the WFC were 13,172. The relative proportion of types and tokens was 2.37% or 1:29. However, the study of vocabulary in the Weather Forecast Corpus (WFC) used only 268,600 tokens (593 lemmas) or 48.33 % of the whole corpus selected from samples in the study. The samples were repeated more than 55 times. After analyzing the weather texts and performing a collocational analysis based on the evidence of corpus, it is possible to observe some patterns and to look for regularities and phraseology of the language. Caution is required for generalizing WFC words as the size of the WFC is limited and it is a very specialized corpus.

5.2 Implications

5.2.1 Pedagogical Implications

5.2.1.1 Implications for Teachers/ Course and Material Designers

This study can show how the results of using corpus-based analysis can be relevant to 'Teaching English for Specific Purposes' (TESP) and to syllabus and materials design. The corpus-based analysis can also be a part of regular language classes, improving students' learning skills while providing them with meaningful and authentic language data. The teachers, course and material designers can apply data from the corpus to help develop and evaluate vocabulary for the teaching and learning of the English language. Since it is often the case that when English teachers are preparing exercises or tests for their students, they need to find or invent sentences which focus on specific words or grammar structure. Concordance makes it very easy for the English teachers to locate such structures in the context of authentic sentences, with which they can then use to prepare tests or exercises for their students.

The use of corpus-based analysis to study the proportions of vocabulary in Weather Forecast Corpus texts results in insights for language teachers on the perspectives of types of vocabulary in a specialized text as well as implications for ESP pedagogy. Technical vocabulary, including technical noun phrases should be included in language teaching, especially for meteorologists, and meteorological students who are beginners in ESP classes of Thai Meteorological Institute (TMI). It is a reasonable assumption from all results that this study can

be taught in principle, by teachers unfamiliar with technical terms, to all types of meteorological disciplines in the early stages of meteorological student's career in the future, and that it will prove useful to all of them throughout their studies. The teachers can ensure that the language they present in the courses corresponds to the language that is required by the meteorological students and meteorologists in TMI in their weather forecasting area.

These findings can be applied in English for Academic Purposes (EAP) course for the meteorological students and English for Occupational Purposes (EOP) course for the meteorologists and meteorological officers, to help them improve and strengthen their knowledge of the vocabulary as well as their skill in writing and reading comprehension. For the knowledge of the language, it is significant for English for Academic Purposes (EAP) / English for Specific Purposes (ESP) teachers to know what to include and introduce as part of language teaching, especially technical words or vocabulary (Wasuntarasophit, 2008). Course and material designers should emphasize both technical noun phrases and technical words and put them in the course/ teaching materials for the meteorological students, meteorologists and meteorological officers. The English courses for TMI should contain the following objectives to improve the communicative skills of the meteorological students, meteorologists.

a) To be able to use vocabulary in English weather forecast texts

b) To explain the weather forecast in English

c) To be able to write or read weather forecast texts in English

d) To be aware of the technical single words, technical noun phrases, and standard abbreviations about weather forecast texts

5.2.1.2 Implications for Learners in Thai Meteorological Institute (TMI)

The findings from the corpus show that many technical phrases occurred in the WFC which are problem for the meteorologists and meteorological students (personal communication). This points out that they should pay more attention to learning these noun phrases, not only what they mean but how they are interpreted. Also, they should study both the common meaning as well as the technical meaning according to the weather forecasting context. They have to be aware that some general words from the General Service Lists (GSL) can provide technical meaning when used in weather forecast texts. For meteorologists and meteorological students, if they cannot understand technical meaning in English, they will have problems when they want to be professional meteorological officers or meteorologists. Therefore, they need to prepare themselves for their careers in the future with knowledge of meteorological vocabularies.

5.2.2 Occupational Implications

5.2.2.1 Implications for Meteorologists

The following general templates could be designed for each type of weather forecast to be used by meteorologists who were the beginner of meteorologists and have to write weather forecasts:

Type 1: Warning Advisory (WF1)

- .....are expected in..... for a few days.

Type 2: Daily Forecast (WF2)

- ..... Thailand..... in the ..... with..... in .....
  Minimum temperature ..... °C Maximum temperature .....°C
  ..... winds.....km/hr.

Type 3: Shipping Forecast (WF3)

- ..... prevails over the ..... should proceed with this caution.

Type 4: Weekly Forecast (WF4)

- ..... will ..... Thailand..... should beware of.....
  especially..... on.....

Type 5: Monthly Forecast (WF5)

- This..... season, the amount of rainfall early in the season will be ..... and then .....

The junior meteorologists who have to write weather forecasts can apply the general templates mentioned above for writing weather forecasts in English on their own. They will learn how to write a well-organized weather forecast text after seeing a lot of good and standard examples from the authentic texts.

5.2.2.2 Implications for Meteorological Officers

As the findings in this study were taken from real-life texts, they can be applied to meteorological officers to develop and assess vocabulary for the learning of meteorological language in meteorological instrument's manuals or textbooks. This will hopefully lead to the improve their ability to cope with suitable work in reading weather forecast texts. Eventually it will lead to the development of English used at the meteorological offices as a whole.

5.3 Suggestions for Further Studies

This research is only a tentative small scale study on the vocabulary used in weather forecast texts. The field of weather forecasting (synoptic meteorology) is a field of meteorology. Therefore, further studies should include in many other fields in meteorology such as aeronautical meteorology, agricultural meteorology, hydrological meteorology, maritime meteorology, biometeorology, astrometeorology, military meteorology, medical meteorology, or highways meteorology, etc.

In terms of vocabulary analysis, to get more actual proportions of vocabulary categories, single words and multi-word items should be analyzed separately, and both of them should be identified by checking against not only the general vocabulary and the academic vocabulary but also their meaning according to the context they are in. Low frequent words should not be ignored in the specialized corpus since sometimes low frequent words may be technical terms. The researcher needs to be careful in separating single words from a phrase because it would impair the accuracy of the real meaning of the multiword lexical units when used in the phrase.

Dictionary of meteorological terms should be compiled using corpus-based analysis to select representative headwords. Further studies should also focus on meteorological symbols both letters and number. Multiword units in meteorological texts should also be carried out. Thus, larger scale of meteorological corpus is needed. Designing a corpus-based (English and Thai) meteorological dictionary will better carter for the needs of users especially meteorologists, meteorological officers, English teachers and meteorological students in Thailand.

BIBLIOGRAPHY

American Meteorological Society. 2000. **Glossary of Meteorology.** [Online]. Available :http://amsglossary.allenpress.com/glossary.

Aroonmanakun, W. 2005. "Collocation Extract: A Tool for Extracting Collocation." **Journal of English Studies.** 2 : 28-39.

Bahns, J. and Eldaw, M. 1993. "Should We Teach EFL Students Collocations?." **System.** 21 (1) : 101-114.

Baker, M. 1992. **In other Words.** London : Routledge.

Bangkok Post. 2008. **Thailand Weather.** Bangkok : Bangkok Post.

Barnhart R.K. Editor. 1995. **Barnhart Abbreviation Dictionary.** 2nd ed. United States of America : John Wiley and Sons.

Bauman, J. and Culligan, B. 1995. **The General Service List.** [Online]. Available :http://Jbauman.com/gsl.html.

Benson, M. 1990. "Collocations and General-purpose Dictionaries." **International Journal of Lexicography.** 3 : 23-34. [Online]. Avialable :http://www.ijl.oxfordjournals.org/cgi/pdf_extract/5/4/323.

Benson, M. Benson, E . and Ilson, R. 1986a. **The BBI Combination Dictionary of English.** Amsterdam : John Benjamins.

Biber, D. Conrad, S. and Reppen, R. 1998. **Corpus Linguistics: investigating language structure and language use.** United Kingdom : Cambridge University Press.

Boas, H. C. 2005. "Semantic Frames as Interlingual Representations for Multilingual Lexical Databases." **International Journal of Lexicography.** 18(4) : 445-478.

Boonyasaquan, S. 2005. "An Analysis of Collocational Volations in Translation." Humanities Journal. 27(2): 79-91. [Online]. Available :http://ejournals.swu.ac.th/index.php/hm/article/view/618

Bowker, L. and Pearson, J. 2002. Working with Specialized Language: A practical **guide to using corpora.** London : Routledge.

Carthy, M. 1990. **Vocabulary.** East Kilbride : Oxford University Press.

Chung, T. M. and Nation, P. 2003. "Technical Vocabulary in Specialised Texts." **Reading in a Foreign Language.** 15 (2) : 103-116.

Cowan, J. R. 1974. Lexical and Syntactic Research for the Design of EFL Reading Materials. **TESOL Quarterly**. 8(4) : 389-400.

Coxhead, A. 2000. "A New Academic Word List." **TESOL Quarterly**. 34 : 213-238.

Coxhead, A. and Nation, P. 2001. "The Specialized Vocabulary of English for Specific Purposes." 252-267. in Flowerdew, J. and Peacock, M. **Research Perspectives on English for Academic Purposes.** 2nd ed. Cambridge : Cambridge University Press.

Curzan, A. and Adams, M. 2006. **How English Works: A linguistic introduction.** United States of America : Pearson Education.

Dunlop, S. 2001. **Dictionary of Weather.** Oxford : Oxford University Press.

Durham, M. 1994. "Linguistic and Design Elements in Australian Computer Manuals: Effects on Legibility and Readability". Ph.D. Thesis: Macquarie University.

Ellendersen, J. 2007. **Grammar in Dictionaries of Languages for Special Purposes**. [Online]. Available :http://pure.au.dk/portal-asb-student/files/1462/000161028-161028.pdf

Farrell, P. 1990. "Vocabulary in ESP: A lexical analysis of the English of electronics and study of semi-technical vocabulary." **CLCS Occasional Paper**. 25 Trinity College.

Flowerdew, J. 1993. "Concordancing as a Tool in Course Design." **System.** 21(2) : 231-244.

Forthofer, R.N. Lee, E.S. and Hernandez, M. 2007. Biostatistics : A guide to design, analysis, and discovery. 2nd ed. United States of America : Elsevier.

Fuertes, O. and Pedro, A. 2007. "A Corpus-based View of Lexical Gender in Written Business English". **English for Specific Purposes**. 26 (4) : 219-234.

Garrigues, S. 1999. **Overcoming Pronunciation Problems of English Teacher in Asia.** [Online]. Available :http://asianbridges.com/pac2/presentations/garrigues.html.

Goulden, R., Nation, P., and Read, J. 1990. "How large can a receptive vocabulary be?". **Applied Linguistics**. 11(4) : 341-363.

Greenbaum, S. 1996. **The Oxford English Grammar.** Oxford : Oxford University Press.

Hartmann, R.R.K. and James, G. 1998. **Dictionary of Lexicography.** United Kingdom: Routledge.

Hartmann, R.R.K. 2003. **Lexicography: Critical Concepts.** 4th ed. United Kingdom: Routledge.

Healey, J. 1993. **Statistics: A Tool for Social Research**. Belmont: Wadsworth.

Herbst, T. 1996. "What are collocations: Sandy Beaches or False Teeth?." **English Studies.** 4 : 379-393.

Hockey, S.M. 2002. **Electronic Texts in the Humanities: Principles and Practice.** United States of American: Oxford University Press.

Hunston, S. 2002. **Corpora in Applied Linguistics.** Cambridge : Cambridge University Press.

James, C. 1998. **Errors in Language Learning and Use: Exploring error analysis.** New York: Addison Wesley Longman.

Jurafsky, D. and Martin, J.H. 2002. **Speech and Language Processing: An introduction to natural language processing, computational linguistics and speech recognition.** NJ: Prentice-Hall.

Kaewphanngam, C. 2002. "A Corpus Analysis of Psychology Texts as a Basis for the Development of Teaching Materials in English for Academic Purposes." M.A. Thesis, Mahidol University.

Kaewphanngam, C. Broughton, M.M. and Soranasataporn, S. 2002. "Corpus-Based Analysis: Guidelines for Getting Practical Language Input in Materials Development." **SLLT**. 11 : 16-32. [Online]. Available: http://www.sc.mahidol.ac.th/sclg/sllt/ArtileSLLT_Maurice_.pdf.

Kennedy, G. 1998. **An Introduction to Corpus Linguistics**. Harlow : Peason Education.

Kjellmer, G. 1987. **Aspects of English Collocations**. Amsterdam : Rodopi.

Krishnamurthy, R. 1987. "The Process of Compilation." in Sinclair, J.M. **Looking Up: An account of the COBUILD project in lexical computing ELT.** 2nd ed. London: Collins.

Krishnamurthy, R. 2000. **Classroom Cornucopia : The COBUILD dictionary and the Bank of English Corpus.** [Online]. Available :http://www.cobuild.collins.co.th.

Landau, B. 2001. **Perceptual units and their mapping with language**. New York : Elsevier.

Lewis, M. 2000. "Language in the Lexical Approach." 126-154. in Lewis, M. **Teaching Collocation : Further Development in the Lexical Approach** 2nd ed. Oxford : Oxford University Press.

Leech, G. 1997. "Teaching and language corpora: a convergence." in Candlin, C.N. **Teaching and Language Corpora.** 3rd ed. New York : Longman.

Leech, G. Rayson, P. and Wilson, A. 2001. **Word Frequencies in Written and Spoken English.** 2nd ed. Malaysia: Pearson Education.

Macmillan, 2009. **How was the Corpus used in Creating the Dictionary?.** [Online]. Available: http://www.macmillandictionary.com/about.html

Mallikamas, P. 1999. "Application of Corpora in Language Teaching." **Thai TESOL BULLETIN**. 12 : 1-17.

Martin, A.V. 1976. "Teaching Academic Vocabulary to Foreign Graduate Students." **TESOL Quarterly.** 10(1) : 91-97.

Martin, J. R. 2001. "Technicality and abstraction: language for the creation of specialized texts." 211-228. in Burns, A. and Coffin, C. 2nd ed. **Analyzing English in a Global Contexts: a Reader**. London : Routledge.

Meyer, C.F. 2002. **English Corpus Linguistics: An Introduction**. Cambridge : Cambridge University Press.

Maunder, W.J. 1992. **Dictionary of Global Climate Change**. London: UCL Press.

Nation, I.S.P. 2001. **Learning Vocabulary in Another Language.** Cambridge: Cambridge University Press.

Opitz, K. 1983. "Special-Purpose Lexicography: Dictionaries for technical use." 89-95. in Hartmant R.R.K. Editor. **Dictionaries and Their Uesrs, Exeter Linguistic Studies 4.** Exeter : University of Exeter Press.

Oxford University. 2010. **The British National Corpus.** [Online]. Available: http://www.natcorp.ox.ac.uk. Accessed January 26, 2009.

Partington, G. 2005. "The role of Professional Development in Creating a New Role in Schools : Inducting Coordinators into the Follow the Dream' Program." in The Symposium Issues in selecting for success at the Australian Association for Research in Education 2005 International Education Research Conference. The University of Western Sydney, Parramatta : NSW.

Pongnan, P. 2006. "Corpus-based Analysis of English Vocabulary in Pharmaceutical Leaflets and a Design of a Sample Dictionary." M.A. Thesis, King Mongkut's Institute of Technology Ladkrabang.

Potipiti, T. Sornlertlamvanich, V. and Charoenporn. S. 2000. **Toward Building a Corpus-Based Dictionary for Non-Word-Boundary Language.** [Online]. Available: http://conf.org/Irec2000/Korterm.kaist.ac.kr/Irec2000/wtrcfile/TanapongPotipiti.pdf.

Redman, S. and Edwards, L. 2003. **English Vocabulary in Use (pre-intermediate and intermediate).** 2nd ed. United Kingdom : Cambridge University Press.

Renouf, A.J. 1987. "Corpus Development." 1-40. in Sinclair, J. Looking Up : An account of the COBUILD Project in lexical computing. 2nd ed. London: Collins.

Rieter, E. and Sripada, S. 2003. "Learning the Meaning and Usage of Time Phrases from a Parallel Text-Data Corpus." 78-85. in Proceedings of the HLT-NAA CL03 Workshop on Learning Word Meaning from Non-Linguistic Data. [Online]. Available: http://www.aclweb.org/anthology/W/W03/W03-0611.pdf.

Runncie, M. 2002. **Oxford Collocations : Dictionary for Students of English**. Italy: Oxford University Press.

Saslow, J.M. and Mongillo, J.F. 1985. **English in Context : Reading comprehension for science and technology.** United States of America: Prentice-Hall.

Scot, M . 1998. Wordsmith Tools. Oxford : Oxford University Press.

Sievert, T. 2005. **Weather Forecasting**. United States of America : Capstone Press.

Simon, M. 1996. "Linguistics Corpora and Language Teaching." **Annual Review of Applied Linguistics.** 16 : 182-199.

Sinclair, J. 1991. **Corpus, Concordance, Collocation**. Oxford : Oxford University Press.

Sola, R. D. Stahl, D. and Kerchelich, K1995. **Abbreviations Dictionary**. 2nd ed. United States of American : CRC Press.

Sripada, S. et al.. 2002. **Segmenting Time Series for Weather Forecasting**. [Online]. Available: http://www.csd.abdn.ac.uk/vereiter/papers/es2002a.pdf.

Srisirirungrot, M. 1998. "An Analysis of Mass Media Weather Forecast News." M.A. Thesis in Mass Communication Faculty of Communication Arts, Chulalongkorn University.

Thai Meteorological Department. 1979. **Glossary of Meteorological Terms**. Bangkok : Thai Meteorological Department.

Thai Meteorological Department. 2006. **Thai Meteorological Department' s Visions.** [Online]. Available :htpp://www.tmd.go.th/en/aboutus/vision.pp.

Thai Meteorological Department. 2007. **Weather Forecasting and Warning: Weather forecasts.** [Online]. Available :http://www.tmd.go.th/info/infoFile?Id=63.

Thai Meteorological Department. 2007. **Glossary of Meteorological Terms (English-Thai).** [Online]. Available :http://www.tmd.go.th/met_dict.php.

The Royal Institute. 2006. **Geographical Dictionary (English-Thai).** Bangkok : the Royal Institute.

Translation Department at Chulalongkorn. 1994. **Oxford-Duden:Pictorial Thai and English Dictionay.** Bangkok : Amarin.

Trask, R. L. 1999. **Key Concepts in Language and Linguistics.** New York : Routledge.

Tribble, C. and Jones, G. 1990. **Concordance in the Classroom.** London : Longman.

Tuck, A. 1990. Oxford Business English Dictionary. 3rd ed. Oxford : Oxford University Press.

Vitayapirak, J. and Pornchalermpong, P. 2005. **Dictionary of Food Engineering (English-Thai).** Bangkok : Se-education.

Vongpumivitch, V. Ju-yu, H. and Yu-Chia, C. 2009. "Frequency Analysis of the Words in the Academic Word List (AWL) and Non-AWL Content Words in Applied Linguistics Research Papers." English for Specific Purposes. 33-41.

Wasuntarasophit, S. 2008. Technical and Academic Vocabulary in Electrical Engineering Textbooks. Ph.D. Thesis, Suranaree University of Technology.

West, M. 1953. **A General Service List of English Words.** London : Longman.

World Meteorological Organization, 2010. **About us.** [Online]. Available : http://www.wmo.int/pages/about/index_en.html

Zimmerman, O.T. Lavine, I. 1949. **Scientific and Technical Abbreviation, Signs and Symbols.** New Hampshire : Industrial Research Service.

Zughoul, M.R. 1991. Lexical Choice : Towards writing problematic word Lists. International Review of Applied Linguistics in Language Teaching. [Online]. Available : http://EBSCO host Research Databases. (No. 0019042X).

APPENDICES

APPENDIX A

APPENDIX A

The Word Frequency List and Abbreviations of the Weather Forecast Corpus (not lemmatized)

N	= Noun	Pron	= Pronoun	Abbrev	= Abbreviation,
Adj	= Adjective	Det	= Determiner	N(abbrev)	= Noun (abbreviation)
Adv	= Adverb	N(Inf)	= Noun (Infinitive)	Adv(Neg)	= Adverb (Negative)
V	= Verb	PV	= Primary verb		
Prep	= Preposition	MV	= Modal verb		
Conj	= Conjunction	N(Num)	= Noun (Number)	Freq.	= Frequency

Rank	Word	Word Classes	Freq.	%
1.	the	Det.	20,565	3.70
2.	and	Conj	12,577	2.26
3.	of	Prep/Conj	10,025	1.80
4.	in	Prep/Adv	7,956	1.43
5.	to	Inf./Prep	6,434	1.16
6.	°C	Abbrev	4,244	0.76
7.	a	Det	3,969	0.71
8.	with	Prep	3,926	0.71
9.	temperature	N	3,461	0.62
10.	thundershowers	N	3,371	0.61
11.	km	N(abbrev)	3,119	0.56
12.	for	Prep	3,107	0.56
13.	is	Pv	2,869	0.52
14.	Hr.	N(abbrev)	2,765	0.50
15.	Thailand	N	2,724	0.49
16.	scattered	Adj	2,702	0.49
17.	at	Prep	2,439	0.44
18.	from	Prep	2,384	0.43
19.	wave	N	2,259	0.41
20.	or	Conj	2,170	0.39
21	wind	N	2,145	0.39
22	height	N	2,135	0.38
23	during	Prep	2,056	0.37

Rank	Word	Word Classes	Freq.	%
24	weather	N	2,011	0.36
25	meters	N	1,958	0.35
26	isolated	Adj	1,956	0.35
27	on	Prep	1,945	0.35
28	winds	N	1,819	0.33
29	forecast	N	1,746	0.31
30	be	PV	1,734	0.31
31	area	N	1,723	0.31
32	rain	N	1,626	0.29
33	heavy	Adj	1,617	0.29
34	are	PV	1,614	0.29
35	cloudy	Adj	1,605	0.29
36	sea	N	1,571	0.28
37	over	Adv/Prep	1,566	0.28
38	this	Det/Pron/Adv	1,550	0.28
39	that	Det/Pron/Adv/Conj	1,543	0.28
40	part	N	1,478	0.27
41	gulf	N	1,450	0.26
42	minimum	Adj/N	1,357	0.24
43	about	Prep/Adv	1,334	0.24
44	maximum	Adj/N	1,301	0.23
45	will	MV	1,247	0.22
46	Precip	N(abbrev)	1,239	0.22
47	°F	N(abbrev)	1,212	0.22
48	by	Prep/Adv/Adj	1,204	0.22
49	should	MV	1,175	0.21
50	as	Conj/Prep/Adv	1,169	0.21
51	time	N	1,124	0.20
52	areas	N	1,123	0.20
53	knots	N	1,113	0.20
54	southwesterly	Adj	985	0.18
55	Andaman	N	978	0.18
56	upper	Adj/N	957	0.17
57	widely	Adv	901	0.16

Rank	Word	Word Classes	Freq.	%
58	high	Adj/N	891	0.16
59	northeasterly	Adj	880	0.16
60	was	PV	870	0.16
61	period	N	863	0.16
62	it	Pron	850	0.15
63	pm	N(abbrev)	849	0.15
64	precipitation	N	847	0.15
65	pressure	N	834	0.15
66	average	N	803	0.14
67	Az	N(abbrev)	790	0.14
68	coast	N	788	0.14
69	mostly	Adv	779	0.14
70	A.M./a.m.	N(abbrev)	752	0.14
71	southern	N	751	0.14
72	an	Det	737	0.13
73	actual	Adj	736	0.13
74	m	N(abbrev)	727	0.13
75	UTC	N(abbrev)	727	0.13
76	data	N	725	0.13
77	very	Adv/adj	718	0.13
78	morning	N/Adj	714	0.13
79	east	N/Adj	696	0.13
80	may	N/MV	689	0.12
81	h.m.	Abbrev	687	0.12
82	low	N/Adj	672	0.12
83	Bangkok	N	668	0.12
84	meteorological	Adj	657	0.12
85	rise	V/N	655	0.12
86	no	Adv (Neg)/ N(abbrev)	648	0.12
87	max	N(abbrev)	637	0.11
88	set	V/N	626	0.11
89	all	N/Adj	622	0.11
90	we	Abbrev/Pron	619	0.11

Rank	Word	Word Classes	Freq.	%
91	meter	N	614	0.11
92	some	Adj/Pron/Adv	612	0.11
93	lower	Adj	611	0.11
94	daily	Adj/Adv/N	597	0.11
95	information	N	597	0.11
96	which	Pron/Adj	595	0.11
97	cold	Adj/N	591	0.11
98	central	Adj/N	585	0.11
99	rainfall	N	576	0.10
100	widespread	Adj	562	0.10
101	cool	Adj/N	557	0.10
102	than	Conj/Prep	551	0.10
103	week	N	551	0.10
104	south	N/Adj	543	0.10
105	climate	N	534	0.10
106	fog	N	533	0.10
107	.th	N(abbrev)	533	0.10
108	partly	Adv	521	0.09
109	air	N	514	0.09
110	water	N	508	0.09
111	not	Adv (Neg)	504	0.09
112	more	Adj/Det/N	500	0.09
113	monsoon	N	495	0.09
114	's	N(abbrev)	482	0.09
115	min	N(abbrev)	481	0.09
116	northern	Adj/N	479	0.09
117	west	Adj/N	476	0.09
118	almost	Adv	475	0.09
119	easterly	Adj/Adv	472	0.08
120	waves	N	472	0.08
121	have	PV	469	0.08
122	has	PV	465	0.08
123	can	MV	464	0.08
124	throughout	Prep/Adv	462	0.08

Rank	Word	Word Classes	Freq.	%
125	h	N(abbrev)	461	0.08
126	northeastern	Adj	457	0.08
127	technology	N	450	0.08
128	distribution	N	447	0.08
129	issued	V/Adj	443	0.08
130	communication	N	437	0.08
131	Sa	N(abbrev)	430	0.08
132	N	N(abbrev)	422	0.08
133	were	PV	421	0.08
134	strait	Adj/N	419	0.08
135	malacca	N	417	0.08
136	hot	Adj	416	0.07
137	above	Prep	412	0.07
138	Fr	N(abbrev)	412	0.07
139	eastern	Adj	411	0.07
140	ministry	N	407	0.07
141	northeast	Adj/N	407	0.07
142	Mo	N(abbrev)	401	0.07
143	Mu	N(abbrev)	400	0.07
144	these	Pron/Adj	400	0.07
145	system	N	399	0.07
146	Tu	N(abbrev)	399	0.07
147	sun	N	397	0.07
148	situation	N	395	0.07
149	people	N	393	0.07
150	observations	N	386	0.07
151	LST	N(abbrev)	384	0.07
152	its	Pron	379	0.07
153	used	V/Adj	379	0.07
154	January	N	375	0.07
155	coastal	Adj	371	0.07
156	moderate	Adj	368	0.07
157	model	N	365	0.07
158	synoptic	N	365	0.07

Rank	Word	Word Classes	Freq.	%
159	also	Adv	361	0.06
160	quality	N	360	0.06
161	north	N/Adj	357	0.06
162	normal	N/Adj	354	0.06
163	surface	N	352	0.06
164	hour	N	350	0.06
165	China	N(abbrev)	346	0.06
166	nne	N	346	0.06
167	December	N	339	0.06
168	August	N	337	0.06
169	such	Adj	336	0.06
170	April	N	336	0.06
171	message	N	334	0.06
172	change	N/V	334	0.06
173	mm	N(abbrev)	332	0.06
174	but	Conj/Prep/Adv	329	0.06
175	CFSR	N(abbrev)	329	0.06
176	E	N(abbrev)	328	0.06
177	expected	V/Adj	328	0.06
178	NE	N(abbrev)	325	0.06
179	route	N/V	324	0.06
180	broadcast	N/V	323	0.06
181	cloud	N	323	0.06
182	shipping	N	323	0.06
183	index	N	320	0.06
184	Kotabharu	N	319	0.06
185	Singapore	N	319	0.06
186	when	Adv/Conj/Pron	318	0.06
187	day	N	314	0.06
188	prevails	V	313	0.06
189	southwest	N/Adj/Adv	312	0.06
190	November	N	311	0.06
191	caution	N	310	0.06
192	mean	N/V	309	0.06

Rank	Word	Word Classes	Freq.	%
193	likely	Adj/Adv	308	0.06
194	storm	N	298	0.05
195	metropolis	N	297	0.05
196	moon	N	294	0.05
197	other	Adj/Pron	292	0.05
198	between	Prep/Adv	291	0.05
199	department	N	290	0.05
200	gust	N	290	0.05
201	proceed	V	290	0.05
202	their	Adj	290	0.05
203	next	Adj	286	0.05
204	southeasterly	Adj/Adv	281	0.05
205	into	Prep	280	0.05
206	level	N/Adj	280	0.05
207	tropical	Adj	280	0.05
208	global	Adj	279	0.05
209	while	Conj	278	0.05
210	been	PV	277	0.05
211	February	N	275	0.05
212	rains	N	272	0.05
213	analysis	N	268	0.05
214	said	V	264	0.05
215	across	Prep/Adv	257	0.05
216	one	N(Num)	257	0.05
217	new	Adj	253	0.05
218	long	Adj/ N(abbrev)	250	0.04
219	if	Conj	249	0.04
220	Friday	N	248	0.04
221	ships	N	247	0.04
222	Thursday	N	247	0.04
223	September	N	245	0.04
224	after	Adj/Prep/Adv/Conj	243	0.04
225	et	N(abbrev)	243	0.04
226	Jul	N(abbrev)	241	0.04

Rank	Word	Word Classes	Freq.	%
227	research	N/V	241	0.04
228	days	N	240	0.04
229	beware	V	239	0.04
230	conditions	N	239	0.04
231	they	Pron	236	0.04
232	atmospheric	Adj	234	0.04
233	small	Adj	233	0.04
234	J	N(abbrev)	232	0.04
235	southerly	Adj/Adv	232	0.04
236	below	Adv	230	0.04
237	forecasts	N/V	230	0.04
238	station	N	229	0.04
239	places	N/V	227	0.04
240	barometer	N	224	0.04
241	humidity	N	224	0.04
242	Jun	N(abbrev)	224	0.04
243	metres	N	224	0.04
244	using	N/V	224	0.04
245	first	N(Num)	222	0.04
Rank	word	Word Classes	Freq	%
246	Wednesday	N	222	0.04
247	vicinity	N	220	0.04
248	along	Prep/Adv	219	0.04
249	there	Adv/Pron	219	0.04
250	strong	Adj	216	0.04
251	two	N(Num)	215	0.04
252	through	Prep/Adv	211	0.04
253	light	N/Adj	210	0.04
254	national	Adj/N	210	0.04
255	Tuesday	N	210	0.04
256	ice	N	209	0.04
257	October	N	209	0.04
258	most	Adj/Pron/N	208	0.04
259	hPa/HPA	N(abbrev)	207	0.04

Rank	Word	Word Classes	Freq.	%
260	month	N	207	0.04
261	ocean	N	207	0.04
262	based	V/Adj	206	0.04
263	sunny	Adj	206	0.04
264	I	Pron/N(abbrev)	203	0.04
265	where	Adv/Conj/N	202	0.04
266	al.	N(abbrev)	200	0.04
267	both	Adj/Conj	200	0.04
268	general	Adj	197	0.04
269	use	V/N	197	0.04
270	year	N	197	0.04
271	up	Adv/Prep/Adv	196	0.04
272	would	MV	194	0.03
273	fig	N(abbrev)	193	0.03
274	h.t.	N(abbrev)	193	0.03
275	today	N	193	0.03
276	dioxide	N	192	0.03
277	near	Adj/Adv/V	191	0.03
278	radar	N	188	0.03
279	Sunday	N	188	0.03
280	Suratthani	N	188	0.03
281	b	N(abbrev)	185	0.03
282	many	Det/Pron/N/Adj	185	0.03
283	monthly	Adj/Adv	185	0.03
284	because	V	184	0.03
285	drop	N/V	183	0.03
286	flooding	N	183	0.03
287	scale	N	183	0.03
288	measurement	N	182	0.03
289	snow	N	182	0.03
290	western	Adj	182	0.03
291	each	Adv/Pron	180	0.03
292	Monday	N	180	0.03
293	windy	Adj	180	0.03

Rank	Word	Word Classes	Freq.	%
294	could	MV	178	0.03
295	mercury	N	178	0.03
296	systems	N	178	0.03
297	unhealthy	Adj	178	0.03
298	range	N	177	0.03
299	satellite	N	177	0.03
300	so	Adv/Conj	177	0.03
301	available	Adj	175	0.03
302	date	N	175	0.03
303	fall	N/V	175	0.03
304	increase	V/N	175	0.03
305	only	Adv/Adj/Conj	175	0.03
306	science	N	174	0.03
307	cell	N	173	0.03
308	instruments	N	173	0.03
309	university	N	173	0.03
310	clouds	N	171	0.03
311	variable	Adj/N	171	0.03
312	standard	Adj/N	170	0.03
313	amount	N/V	169	0.03
314	www.	N(abbrev)	169	0.03
315	sets	N	168	0.03
316	hurricane	N	167	0.03
317	Lat/lat	N(abbrev)	165	0.03
318	models	N	165	0.03
319	then	Adv/Adj/N	165	0.03
320	well	Adv	165	0.03
321	cover	V	163	0.03
322	showers	N	163	0.03
323	thunderstorm	N	163	0.03
324	mountain	N	162	0.03
325	region	N	161	0.03
326	Saturday	N	161	0.03
327	slight	Adj	161	0.03

Rank	Word	Word Classes	Freq.	%
328	gusty	Adj	160	0.03
329	measurements	N	160	0.03
330	different	Adj	159	0.03
331	NOAA	N(abbrev)	159	0.03
332	rather	Adv	159	0.03
333	however	Conj/Adv	158	0.03
334	operational	Adj	158	0.03
335	since	Prep/Conj	158	0.03
336	WMO	N(abbrev)	158	0.03
337	heat	N	157	0.03
338	warning	N	157	0.03
339	center	N	156	0.03
340	especially	Adv	155	0.03
341	instrument	N	155	0.03
342	large	Adj	155	0.03
343	good	Adj	154	0.03
344	less	Adj	154	0.03
345	't	N(abbrev)	154	0.03
346	error	N	153	0.03
347	made	V	153	0.03
348	'd	N(abbrev)	152	0.03
349	those	Adj/Pron	151	0.03
350	thundershower	N	151	0.03
351	bureau	N	150	0.03
352	must	MV	150	0.03
353	off	Prep/Adv	150	0.03
354	avg	N(abbrev)	149	0.03
355	outlook	N	149	0.03
356	storms	N	149	0.03
357	KPH/kph	N(abbrev)	148	0.03
358	prediction	N	148	0.03
359	trough	N	148	0.03
360	atmosphere	N	147	0.03
361	Chumphon	N	147	0.03

Rank	Word	Word Classes	Freq.	%
362	go.	V/ N(abbrev)	147	0.03
363	much	Adj/Adv	147	0.03
364	speed	N/V	147	0.03
365	active	Adj	146	0.03
366	march	N	146	0.03
367	ozone	N	146	0.03
368	barometers	N	145	0.03
369	radiation	N	145	0.03
370	until	Con/Prep	145	0.03
371	effect	V	144	0.03
372	occur	V	144	0.03
373	out	Adj/Adv/Prep	144	0.03
374	current	N/Adj	143	0.03
375	true	Adj/Adv	143	0.03
376	afternoon	N	141	0.03
377	NCEP	N(abbrev)	141	0.03
378	percent	N/Adj	141	0.03
379	province	N	141	0.03
380	how	Adv	140	0.03
381	land	N	140	0.03
382	temperatures	N	140	0.03
383	oil	N	139	0.03
384	keep	V	137	0.02
385	K	N(abbrev)	136	0.02
386	tel	N(abbrev)	136	0.02
387	portion	N	135	0.02
388	cm	N(abbrev)	134	0.02
389	state	N/V	134	0.02
390	any	Pron/Adj	133	0.02
391	dust	N	133	0.02
392	errors	N	133	0.02
393	resolution	N	133	0.02
394	degree	N	132	0.02
395	three	N(Num)	132	0.02

Rank	Word	Word Classes	Freq.	%
396	world	N	132	0.02
397	country	N	131	0.02
398	zone	N	131	0.02
399	assimilation	N	130	0.02
400	carbon	N	130	0.02
401	changes	N/V	130	0.02
402	offshore	Adj/Adv	130	0.02
403	Chiangmai	N	129	0.02
404	parts	N	129	0.02
405	including	Prep	128	0.02
406	possible	N/Adj	128	0.02
407	us	Pron/ N(abbrev)	128	0.02
408	within	Prep/Adv	128	0.02
409	cause	N/V	127	0.02
410	found	V	127	0.02
411	observed	V/Adj	127	0.02
412	southward	Adv/N	127	0.02
413	earth	N	126	0.02
414	emissions	N	126	0.02
415	July	N	126	0.02
416	value	N/V	126	0.02
417	June	N	125	0.02
418	moonrise	N	125	0.02
419	moonset	N	125	0.02
420	results	N/V	125	0.02
421	per	N	123	0.02
422	Phuket	N	123	0.02
423	example	N	121	0.02
424	thunderstorms	N	121	0.02
425	international	Adj/N	120	0.02
426	vertical	Adj/N	120	0.02
427	American	Adj	119	0.02
428	ensemble	N	119	0.02
429	following	Adj	119	0.02

Rank	Word	Word Classes	Freq.	%
430	R	N(abbrev)	119	0.02
431	relative	N	119	0.02
432	report	N	119	0.02
433	smaller	Adj	119	0.02
434	thermometer	N	119	0.02
435	provinces	N	118	0.02
436	same	Adj/Pron	118	0.02
437	Nakhonratchasima	N	116	0.02
438	see	V	116	0.02
439	still	Adj/Adv	116	0.02
440	sunset	N	116	0.02
441	before	Prep/Conj/Adv	115	0.02
442	given	V	115	0.02
443	number	N	115	0.02
444	shore	N/Adj	115	0.02
445	valid	Adj	114	0.02
446	result	N/V	113	0.02
447	shown	V	113	0.02
448	ashore	Adv	112	0.02
449	rises	V/N	112	0.02
450	sunrise	N	112	0.02
451	being	N	110	0.02
452	events	N	110	0.02
453	risky	Adj	110	0.02
454	visibility	N	110	0.02
455	end	N/V	109	0.02
456	society	N	109	0.02
457	values	N/V	109	0.02
458	warming	N	109	0.02
459	calculated	V/Adj	108	0.02
460	effects	V/N	108	0.02
461	local	Adj	108	0.02
462	soil	N	108	0.02
463	season	N	107	0.02

Rank	Word	Word Classes	Freq.	%
464	several	Adj	107	0.02
465	space	N	107	0.02
466	united	Adj	107	0.02
467	field	N	106	0.02
468	forecasting	N	106	0.02
469	important	Adj	106	0.02
470	Pacific	N	106	0.02
471	azimuth	N	105	0.02
472	base	N/V	105	0.02
473	calibration	N	105	0.02
474	decrease	V/N	105	0.02
475	evening	N	105	0.02
476	falls	V/N	105	0.02
477	provide	V	105	0.02
478	study	N/V	105	0.02
479	uncertainty	N	105	0.02
480	clear	V/Adj	104	0.02
481	make	V	104	0.02
482	severe	Adj	104	0.02
483	years	N	104	0.02
484	environmental	Adj	103	0.02
485	states	N	103	0.02
486	worst	Adj/Adv/N	103	0.02
487	increasing	N	102	0.02
488	Mexico	N	102	0.02
489	previous	Adj	102	0.02
490	accuracy	N	101	0.02
491	due	N/Adv/Adj	101	0.02
492	further	Adv/Adj/V	101	0.02
493	solar	N	101	0.02
494	Ubonratchathani	N	101	0.02
495	hazardous	Adj	99	0.02
496	observation	N	99	0.02
497	reanalysis	N	99	0.02

Rank	Word	Word Classes	Freq.	%
498	slightly	Adv	99	0.02
499	associated	Adv/V	98	0.02
500	few	Adj/N/Pron	98	0.02
501	plus	Prep/Adj/N	98	0.02
502	policy	N	98	0.02
503	thermometers	N	98	0.02
504	TMD	N(abbrev)	98	0.02
505	Songkhla	N	97	0.02
506	Texas	N	97	0.02
507	W	N(abbrev)	97	0.02
508	case	N	96	0.02
509	late	Adj	96	0.02
510	move	V	96	0.02
511	remark	V/N	96	0.02
512	chapter	N	95	0.02
513	intensifying	N	94	0.02
514	Tak	N	94	0.02
515	taken	V	94	0.02
516	total	N/Adj	94	0.02
517	various	Adj	94	0.02
518	e.g.	N(abbrev)	93	0.02
519	even	Adj/V/Adv	93	0.02
520	ambient	N	92	0.02
521	Chanthaburi	N	92	0.02
522	ground	N	92	0.02
523	least	Adj/Adv	92	0.02
524	service	N	92	0.02
525	development	N	91	0.02
526	every	Adj	91	0.02
527	except	Prep/Conj/V	91	0.02
528	around	Adv/Adj/Prep	90	0.02
529	elsewhere	Adv	90	0.02
530	energy	N	90	0.02
531	legend	N	90	0.02

Rank	Word	Word Classes	Freq.	%
532	microns	N	90	0.02
533	Ranong	N	90	0.02
534	sulphur	N	90	0.02
535	Vietnam	N	90	0.02
536	abstract	N	89	0.02
537	layer	N	89	0.02
538	monoxide	N	89	0.02
539	nitrogen	N	89	0.02
540	point	N/V	89	0.02
541	prevail	V	89	0.02
542	top	N/Adj	89	0.02
543	ad	N(abbrev)	88	0.02
544	causes	V/N	88	0.02
545	fairly	Adv	88	0.02
546	Greenwich	N	88	0.02
547	now	Adv/Conj/N	88	0.02
548	according	Adj	86	0.02
549	Alex	N	86	0.02
550	covers	V	86	0.02
551	elevation	N	86	0.02
552	form	N	86	0.02
553	had	PV	86	0.02
554	related	V/Adj	86	0.02
555	Uthaithani	N	86	0.02
556	Nakhonsithammarat	N	85	0.02
557	our	Adj	85	0.02
558	Khonkean	N	84	0.02
559	last	N/Adv/Adj	84	0.02
560	later	Adv	84	0.02
561	moisture	N	84	0.02
562	Aug	N(abbrev)	83	0.01
563	convective	N	83	0.01
564	initial	Adj	83	0.01
565	keywords	N	83	0.01

Rank	Word	Word Classes	Freq.	%
566	lead	V	83	0.01
567	mountainous	Adj	83	0.01
568	readings	N	83	0.01
569	regional	Adj	83	0.01
570	required	V/Adj	83	0.01
571	section	N	83	0.01
572	stations	N	83	0.01
573	studies	N	83	0.01
574	them	Pron	83	0.01
575	under	Prep/Adv	83	0.01
576	comparison	N	82	0.01
577	convection	N	82	0.01
578	haze	N	82	0.01
579	higher	Adj	82	0.01
580	liquid	N	82	0.01
581	impact	N/V	81	0.01
582	intense	Adj	81	0.01
583	laboratory	N	81	0.01
584	P	N(abbrev)	81	0.01
585	summer	N	81	0.01
586	vortex	N	81	0.01
587	caused	V/Adj	80	0.01
588	director	N	80	0.01
589	Nakhonphanom	N	80	0.01
590	term	N	80	0.01
591	Kanchanaburi	N	79	0.01
592	metre	N	79	0.01
593	requirements	N	79	0.01
594	TIGGE	N(abbrev)	79	0.01
595	frequency	N	78	0.01
596	group	N/V	78	0.01
597	Phangnga	N	78	0.01
598	better	Adj/Adv	77	0.01
599	bring	V	77	0.01

Rank	Word	Word Classes	Freq.	%
600	like	V	77	0.01
601	measuring	N	77	0.01
602	significant	Adj	77	0.01
603	thick	Adj	77	0.01
604	tip	N	77	0.01
605	Trat	N	77	0.01
606	version	N	77	0.01
607	do	PV	76	0.01
608	generally	Adv	76	0.01
609	lies	N/V	76	0.01
610	middle	Adj/N	76	0.01
611	moving	N	76	0.01
612	present	N/V	76	0.01
613	second	N(Num)	76	0.01
614	condition	N	75	0.01
615	emission	N	75	0.01
616	fax	N(abbrev)	75	0.01
617	four	N(Num)	75	0.01
618	future	Adj/N	75	0.01
619	greater	Adj	75	0.01
620	levels	N	75	0.01
621	lightning	N	75	0.01
622	meteorology	N	75	0.01
623	Nan	N	75	0.01
624	periods	N	75	0.01
625	boats	N	74	0.01
626	departure	N	74	0.01
627	depression	N	74	0.01
628	doppler	N	74	0.01
629	gauge	N	74	0.01
630	particular	Adj	74	0.01
631	patches	V/N	74	0.01
632	x	N(abbrev)	74	0.01
633	angle	N	73	0.01

Rank	Word	Word Classes	Freq.	%
634	approach	N/V	73	0.01
635	early	Adv	73	0.01
636	method	N	73	0.01
637	natural	Adj	73	0.01
638	observing	N	73	0.01
639	sciences	N	73	0.01
640	westerly	Adj/N	73	0.01
641	again	Adv	72	0.01
642	Chiang Rai	N	72	0.01
643	depth	N	72	0.01
644	down	N/V/Adv	72	0.01
645	l	N(abbrev)	72	0.01
646	regions	N	72	0.01
647	reported	V/N	72	0.01
648	roadside	N	72	0.01
649	trip	N	72	0.01
650	type	N	72	0.01
651	decreasing	N	71	0.01
652	earthquake	N	71	0.01
653	extend	V	71	0.01
654	knowledge	N	71	0.01
655	known	V	71	0.01
656	mist	N	71	0.01
657	order	N/V	71	0.01
658	real	N/Adj	71	0.01
659	standards	N	71	0.01
660	what	Pron/Adj	71	0.01
661	flash	N/Adj	70	0.01
662	hail	N	70	0.01
663	he	Pron	70	0.01
664	indo	N(abbrev)	70	0.01
665	Indochina	N	70	0.01
666	making	N	70	0.01
667	Nong Khai	N	70	0.01

Rank	Word	Word Classes	Freq.	%
668	numerical	Adj	70	0.01
669	provided	V	70	0.01
670	reference	N	70	0.01
671	ridge	N/V	70	0.01
672	scales	N	70	0.01
673	accumulative	N	69	0.01
674	affected	V	69	0.01
675	analyses	V	69	0.01
676	Mae Hong Son	N	69	0.01
677	mainly	Adv	69	0.01
678	methods	N	69	0.01
679	Mukdahan	N	69	0.01
680	paper	N	69	0.01
681	sources	N	69	0.01
682	specific	Adj	69	0.01
683	who	Pron/ N(abbrev)	69	0.01
684	help	V/N	68	0.01
685	key	N	68	0.01
686	measured	V/Adj	68	0.01
687	performance	N	68	0.01
688	Phetchabun	N	68	0.01
689	Surin	N	68	0.01
690	compared	V/Adj	67	0.01
691	dewpoint	N	67	0.01
692	flood	N	67	0.01
693	MWR	N(abbrev)	67	0.01
694	Sakon Nakhon	N	67	0.01
695	table	N	67	0.01
696	take	V	67	0.01
697	velocity	N	67	0.01
698	dry	Adj/V	66	0.01
699	ENE	N(abbrev)	66	0.01
700	fields	N	66	0.01
701	hurricanes	N	66	0.01

Rank	Word	Word Classes	Freq.	%
702	prevailing	N/V	66	0.01
703	another	Adj/Pron	65	0.01
704	covered	V/Adj	65	0.01
705	direction	N	65	0.01
706	exposure	N	65	0.01
707	far	Adv/Adj	65	0.01
708	Kosamui	N	65	0.01
709	often	Adv	65	0.01
710	Phrae	N	65	0.01
711	process	N/V	65	0.01
712	products	N	65	0.01
713	reading	N	65	0.01
714	response	N	65	0.01
715	scientific	Adj	65	0.01
716	scientists	N	65	0.01
717	ship	N	65	0.01
718	therefore	Adv	65	0.01
719	airport	N	64	0.01
720	Alantic	N	64	0.01
721	chance	N	64	0.01
722	extreme	Adj/N	64	0.01
723	NASA	N(abbrev)	64	0.01
724	Oklahoma	N	64	0.01
725	Pattaya	N	64	0.01
726	problem	N	64	0.01
727	Thai	Adj/N	64	0.01
728	types	N	64	0.01
729	applied	Adj/V	63	0.01
730	floods	N	63	0.01
731	gas	N	63	0.01
732	include	V	63	0.01
733	increased	V/Adj	63	0.01
734	particularly	Adv	63	0.01
735	poor	Adj/N	63	0.01

Rank	Word	Word Classes	Freq.	%
736	quantity	N	63	0.01
737	view	N	63	0.01
738	defined	V/Adj	62	0.01
739	disaster	N	62	0.01
740	supercell	N	62	0.01
741	times	N	62	0.01
742	although	Conj	61	0.01
743	daytime	N	61	0.01
744	evidence	N	61	0.01
745	Jan	N(abbrev)	61	0.01
746	miles	N	61	0.01
747	project	N	61	0.01
748	spatial	Adj	61	0.01
749	sum	N/V	61	0.01
750	thus	Adv	61	0.01
751	variables	N	61	0.01
752	comparisons	N	60	0.01
753	flow	N/V	60	0.01
754	improve	V	60	0.01
755	reports	N/V	60	0.01
756	risk	V	60	0.01
757	soc.	N(abbrev)	60	0.01
758	Trang	N	60	0.01
759	CFS	N(abbrev)	59	0.01
760	ECMWF	N(abbrev)	59	0.01
761	fees	N	59	0.01
762	just	Adv/Adj	59	0.01
763	landfall	N	59	0.01
764	office	N	59	0.01
765	potential	N/Adj	59	0.01
766	Satun	N	59	0.01
767	site	N	59	0.01
768	without	Prep/Conj	59	0.01
769	boat	N	58	0.01

Rank	Word	Word Classes	Freq.	%
770	environment	N	58	0.01
771	necessary	Adj/N	58	0.01
772	show	V	58	0.01
773	typhoon	N	58	0.01
774	working	N	58	0.01
775	addition	N	57	0.01
776	back	V	57	0.01
777	climatological	Adj	57	0.01
778	constant	Adj	57	0.01
779	differences	N	57	0.01
780	highest	Adj	57	0.01
781	horizontal	Adj	57	0.01
782	Loei	N	57	0.01
783	online	N/Adv	57	0.01
784	produced	V	57	0.01
785	river	N	57	0.01
786	tops	V/N	57	0.01
787	weakening	N	57	0.01
788	yet	Adv/Conj	57	0.01
789	affect	V	56	0.01
790	assimilated	V	56	0.01
791	best	Adj	56	0.01
792	Buriram	N	56	0.01
793	datasets	N	56	0.01
794	half	Adv/Adj	56	0.01
795	larger	Adj	56	0.01
796	meteor	N	56	0.01
797	northward	Adj	56	0.01
798	place	V/N	56	0.01
799	relatively	Adv	56	0.01
800	torrential	Adj	56	0.01
801	usually	Adv	56	0.01
		Total	**298,599**	**53.36**

APPENDIX B

APPENDIX B

The General Service List (GSL) Found in the Weather Forecast Corpus (Token)

No.	Word	Freq.	Percent	Token
1	the	20,565	3.70	
2	and	12,577	2.26	
3	of	10,025	1.80	
4	in	7,956	1.43	
5	be	7,670	1.38	be (1,618), is (2,869), are (1,614), was (870), were (421), been (277), am (1)
6	to	6,434	1.16	
7	a	4,619	0.83	a (3,882), an (737)
8	wind	3,964	0.71	wind (2,145), winds (1,819)
9	with	3,926	0.71	
10	temperature	3,601	0.65	temperature (3,461), temperatures (140)
11	for	3,107	0.56	
12	wave	2,731	0.49	wave (2,259), waves (472)
13	scatter	2,723	0.49	scattered (2,702), scattering (21)
14	at	2,439	0.44	
15	from	2,384	0.43	
16	or	2,170	0.39	
17	height	2,157	0.39	height (2,135), heights (22)
18	part	2,128	0.38	part (1,478), parts (129), partly (521)
19	during	2,056	0.37	
20	weather	2,011	0.36	
21	on	1,945	0.35	
22	rain	1,898	0.29	rain (1,626), rains (272)
23	heavy	1,617	0.29	
24	sea	1,598	0.29	sea (1,571), seas (27)
25	over	1,566	0.28	
26	this	1,550	0.28	
27	that	1,543	0.28	
28	south	1,526	0.27	south (543), southern (751), southerly (232)
29	about	1,334	0.24	

No.	Word	Freq.	Percent	Token
30	low	1,329	0.24	low (672), lows (10), lower (611), lowest (36)
31	will	1,247	0.22	
32	it	1,229	0.22	it (850), its (379)
33	by	1,204	0.22	
34	time	1,186	0.21	time (1,124), times (62)
35	should	1,175	0.21	
36	as	1,169	0.21	
37	coast	1,169	0.21	coast (788), coasts (10), coastal (371)
38	east	1,168	0.21	east (696), easterly (472)
39	high	1,051	0.19	high (891),highs (21), higher (82), highest (57)
40	have	1,020	0.18	have (469), has (465), had (86)
41	most	987	0.18	most (208), mostly (779)
42	upper	957	0.17	
43	wide	946	0.17	wide (45), widely (901)
44	high	912	0.16	high (891), highs (21)
45	center	861	0.15	center(156), centers (50), centre (54), centres (16), central (585)
46	pressure	845	0.15	pressure (834), pressures (11)
47	use	841	0.15	use (197), uses (41), used (379),using (224)
48	average	811	0.15	average (803), averages (8)
49	set	810	0.15	set (626), sets (168), setting (16)
50	rise	790	0.14	rise (655), rises (112), rising (23)
51	actual	736	0.13	
52	very	718	0.13	
53	morning	714	0.13	
54	may	689	0.12	
55	low	672	0.12	
56	ship	635	0.11	ship (65), shipping (323), ships (247)
57	all	622	0.11	
58	some	612	0.11	
59	week	598	0.11	week (554), weeks (16), weekly (28)
60	daily	597	0.11	
61	which	595	0.11	
62	cold	591	0.11	
63	system	577	0.10	system (399), systems (178)

No.	Word	Freq.	Percent	Token
64	measure	575	0.10	measure (48), measures (40), measured (68), measuring (77), measurement (182), measurements (160)
65	cool	557	0.10	
66	day	554	0.10	day (314), days (240)
67	than	551	0.10	
68	west	549	0.10	west (476), westerly (73)
69	air	514	0.09	
70	water	508	0.09	
71	not	504	0.09	
72	change	501	0.08	change (334), changes (130), changed (13), changing (24)
73	more	500	0.09	
74	cloud	494	0.09	cloud (323), clouds (171)
75	almost	475	0.09	
76	can	464	0.08	
77	storm	447	0.08	storm (298), storms (149)
78	month	417	0.07	month (207), months (25), monthly (185)
79	hot	416	0.07	
80	situation	416	0.07	situation (395), situations (11)
81	north	413	0.07	north (357), northward (56)
82	above	412	0.07	
83	these	400	0.07	
84	hour	397	0.07	hour (350), hours (47)
85	sun	397	0.07	
86	people	393	0.07	
87	surface	386	0.07	surface (352), surfaces (34)
88	like	385	0.07	like (77), likely (308)
89	increase	375	0.07	increase (175), increases (35), increasing (102), increased (63)
90	moderate	371	0.07	moderate (368), moderated (3)
91	cover	367	0.07	cover (163), covers (86), covered (65), covering (53)
92	mean	363	0.07	mean (309), means (47), meant (1), meaning (6)
93	also	361	0.06	
94	quality	361	0.06	quality (360), qualities (1)
95	small	358	0.06	small (234), smaller (119), smallest (5)
96	level	355	0.06	level (280), levels (75)

No.	Word	Freq.	Percent	Token
97	expect	345	0.06	expect (13), expects (4), expected (328)
98	message	344	0.06	message (336), messages (8)
99	cause	338	0.06	cause (127), causes (88), caused (80), causing (43)
100	such	337	0.06	
101	other	336	0.06	other (292), others (44)
102	but	332	0.06	
103	instrument	328	0.06	instrument (155), instruments (173)
104	make	327	0.06	make (104), made (153), making (70)
105	broadcast	325	0.06	broadcast (323), broadcasting (2)
106	say	323	0.06	say (19), says (33), said (264), saying (7)
107	flood	321	0.06	flood (67), floods (63), flooded (8), flooding (183)
108	when	318	0.06	
109	condition	314	0.06	condition (75), conditions (239)
110	base	312	0.06	base (105), bases (1), based (206)
111	station	312	0.06	station (229), stations (83)
112	caution	310	0.06	
113	likely	308	0.06	
114	fall	306	0.05	fall (175), falls (105), fell (8), fallen (4), falling (14)
115	year	301	0.05	year (197), years (104)
116	department	294	0.05	department (290), departments (4)
117	moon	294	0.05	
118	between	291	0.05	
119	next	286	0.05	
120	new	286	0.05	new (253), news (33)
121	place	283	0.05	place (56), places (227)
122	into	280	0.05	
123	while	278	0.05	
124	include	276	0.05	include (63), includes (43), included (42), including (128)
125	general	273	0.05	general (197), generally (76)
126	report	267	0.05	report (119), reports (60), reporting (16), reported (72)
127	show	260	0.05	show (58), shows (54), showed (26), shown (113), showing (9)
128	slight	260	0.05	slight (161), slightly (99)
129	across	257	0.05	
130	one	257	0.05	

No.	Word	Freq.	Percent	Token
131	effect	254	0.05	effect (144), effects (108), effected (2)
132	scale	253	0.05	scale (183), scales (70)
133	if	249	0.04	
134	science	247	0.04	science (174), sciences (73)
135	move	244	0.04	move (96), moves (23), moved (34), moving (76), movement (13), movements (2)
136	after	243	0.04	
137	compare	243	0.04	compare (14), compares (6), compared (67), comparing (14), comparison (82), comparisons (60)
138	provide	243	0.04	provide (105), provides (37), provided (70), providing (31)
139	standard	241	0.04	standard (170), standards (71)
140	result	238	0.04	result (113), results (125)
141	nation	237	0.04	nation (3), nations (18), nation's (6), national (210)
142	state	237	0.04	state (134), states (103)
143	they	236	0.04	
144	value	235	0.04	value (126), values (109)
145	degree	232	0.04	degree (132), degrees (43)
146	ocean	232	0.04	ocean (207), oceans (25)
147	below	230	0.04	
148	read	227	0.04	read (46), red (33), reading (65), readings (83)
149	light	226	0.04	light (210), lights (8), lighting (8)
150	first	222	0.04	
151	large	222	0.04	large (155), larger (56), largest (11)
152	we	222	0.04	
153	along	219	0.04	
154	there	219	0.04	
155	strong	216	0.04	
156	take	214	0.04	take (67), takes (21), took (14), taken (94), taking (18)
157	drop	211	0.04	drop (183), drops (21), dropping (7)
158	through	211	0.04	
159	ice	209	0.04	
160	develop	205	0.04	develop (33), develops (5), developed (50), developing (21), development (91), developments (5)
161	give	205	0.04	give (41), gives (31), gave (11), given (115), giving (7)

No.	Word	Freq.	Percent	Token
162	study	205	0.04	study (105), studies (83), studied (8), studying (9)
163	observe	204	0.04	observe (2), observed (127), observes (2), observing (73)
164	I	203	0.04	
165	health	203	0.03	health (25), unhealthy (178)
166	decrease	202	0.03	decrease (105), decreases (8), decreased (18), decreasing (71)
167	where	202	0.03	
168	both	200	0.03	
169	keep	198	0.03	keep (137), keeps (19), kept (37), keeping (5)
170	general	197	0.03	
171	up	196	0.03	
172	would	194	0.03	
173	today	193	0.03	
174	amount	193	0.03	amount (169), amounts (24)
175	late	192	0.03	late (96), later (84), latest (12)
176	risk	192	0.03	risk (60), risks (22), risky (110)
177	near	191	0.03	
178	many	185	0.03	
179	because	184	0.03	
180	country	182	0.03	country (131), countries (51)
181	snow	182	0.03	
182	western	182	0.03	
183	solution	181	0.03	solution (27), solutions (16), resolution (133), resolutions (5)
184	date	180	0.03	date (175), dates (5)
185	each	180	0.03	
186	find	179	0.03	find (18), finds (3), found (127), finding (12), findings (19)
187	could	178	0.03	
188	operate	178	0.03	operate (158), operates (2), operated (8), operating (10), operational (158)
189	mountain	177	0.03	mountain (162), mountains (15)
190	so	177	0.03	
191	relative	176	0.03	relative (119), relatives (1), relatively (56)
192	speed	176	0.03	speed (147), speeds (29)
193	follow	176	0.03	follow (9), follows (31), followed (15), following (119), followings (2)

No.	Word	Freq.	Percent	Token
194	degree	175	0.03	degree (132), degrees (43)
195	do	175	0.03	do (76), does (51), did (14), done (34)
196	university	174	0.03	university (173), universities (1)
197	lead	174	0.03	lead (83), leads (23), led (25), leading (43)
198	current	172	0.03	current (143), currents (5), currently (24)
199	field	172	0.03	field (106), fields (66)
200	warn	171	0.03	warn (3), warned (11), warning (157)
201	form	170	0.03	form (86), forms (49), formed (21), forming (14)
202	season	170	0.03	season (107), seasons (15), seasonal (48)
203	see	169	0.03	see (116), saw (2), seen (48), seeing (3)
204	warm	166	0.03	warm (52), warms (4), warmed (1), warming (109)
205	shower	166	0.03	shower (3), showers (163)
206	then	165	0.03	
207	well	165	0.03	
208	different	159	0.03	
209	heat	159	0.03	heat (157), heats (2)
210	rather	159	0.03	
211	however	158	0.03	
212	since	158	0.03	
213	especially	155	0.03	
214	good	154	0.03	
215	less	154	0.03	
216	event	153	0.03	event (43), events (110)
217	those	151	0.03	
218	must	150	0.03	
219	off	150	0.03	
220	much	147	0.03	
221	active	146	0.03	
222	march	146	0.03	
223	top	146	0.03	top (89), tops (57)
224	case	145	0.03	case (96), cases (49)
225	until	145	0.03	
226	out	144	0.03	

No.	Word	Freq.	Percent	Token
227	land	143	0.03	land (140), lands (3)
228	true	143	0.03	.
229	point	143	0.03	point (89), points (48), pointed (4), pointing (2)
230	afternoon	141	0.03	
231	how	140	0.03	
232	certain	139	0.03	certaint (29), certainty (5), uncertainty (105)
233	oil	139	0.03	
234	particular	137	0.02	particular (74), particularly (63)
235	example	135	0.02	example (121), examples (14)
236	any	133	0.02	
237	dust	133	0.02	
238	term	133	0.02	term (80), terms (53)
239	world	132	0.02	
240	boat	132	0.02	boat (58), boats (74)
241	present	131	0.02	present (76), presents (6), presented (41), presenting (8)
242	improve	131	0.02	improve (60), improves (8), improved (40), improving (23)
243	type	136	0.02	type (72), types (64)
244	possible	128	0.02	
245	within	128	0.02	
246	extend	127	0.02	extend (71), extends (7), extended (38), extending (11)
247	earth	126	0.02	
248	great	125	0.02	great (34), greater (75), greatest (16)
249	per	123	0.02	
250	lie	123	0.02	lie (47), lies (76)
251	work	123	0.02	work (52), works (8), worked (5), working (58)
252	number	122	0.02	number (115), numbers (7)
253	international	120	0.02	
254	same	118	0.02	
255	calculate	118	0.02	calculate (5), calculated (108), calculating (5)
256	society	116	0.02	society (109), societies (7)
257	still	116	0.02	
258	before	115	0.02	
259	shore	115	0.02	
260	bring	115	0.02	bring (77), brings (5), brought (18), bringing (15)

No.	Word	Freq.	Percent	Token
261	being	110	0.02	
262	end	109	0.02	
263	fair	109	0.02	fair (21), fairly (88)
264	difference	109	0.02	difference (52), differences (57)
265	local	108	0.02	
266	soil	108	0.02	
267	space	108	0.02	space (107), spaces (1)
268	long	107	0.02	
269	several	107	0.02	
270	unite	107	0.02	united (107)
271	important	106	0.02	
272	evening	105	0.02	
273	clear	104	0.02	
274	severe	104	0.02	
275	nature	102	0.02	nature (29), natural (73)
276	remark	102	0.02	remark (96), remarks (5), remarked (1)
277	due	101	0.02	
278	further	101	0.02	
279	group	100	0.02	group (78), groups (22)
280	few	98	0.02	
281	relate	98	0.02	relate (5), relates (2), related (86) relating (5)
282	know	94	0.02	know (20), knows (2), knew (1), known (71)
283	problem	94	0.02	problem (64), problems (30)
284	total	94	0.02	
285	various	94	0.02	
286	even	93	0.02	
287	ground	92	0.02	
288	every	91	0.02	
289	except	91	0.02	
290	around	90	0.02	
291	product	90	0.02	product (25), products (65)
292	elsewhere	90	0.02	
293	frequency	88	0.02	frequency (78), frequencies (10)
294	now	88	0.02	

No.	Word	Freq.	Percent	Token
295	help	87	0.02	help (68), helps (10), helped (5), helping (4)
296	quantity	87	0.02	quantity (63), quantities (24)
297	according	86	0.02	
298	moisture	86	0.02	moisture (84), moistures (2)
299	gas	86	0.02	gas (63), gases (23)
300	last	84	0.02	
301	reference	84	0.02	reference (70), references (14)
302	under	83	0.01	
303	angle	83	0.01	angle (73), angles (10)
304	liquid	82	0.01	
305	paper	82	0.01	paper (69), papers (13)
306	summer	81	0.01	
307	director	80	0.01	
308	scientist	80	0.01	scientist (15), scientists (65)
309	tip	79	0.01	tip (77), tips (2)
310	order	78	0.01	order (71), orders (1), ordered (6)
311	apply	77	0.01	apply (14), applied (63)
312	better	77	0.01	
313	thick	77	0.01	
314	second	76	0.01	
315	table	76	0.01	table (67), tables (9)
316	future	75	0.01	
317	performance	74	0.01	performance (68), performances (6)
318	early	73	0.01	
319	trip	73	0.01	trip (72), trips (1)
320	again	72	0.01	
321	depth	72	0.01	
322	down	72	0.01	
323	direction	72	0.01	direction (65), directions (7)
324	knowledge	71	0.01	
325	real	71	0.01	
326	what	71	0.01	
327	chance	71	0.01	chance (64), chances (7)
328	dry	70	0.01	dry (66), dried (1), drying (3)

No.	Word	Freq.	Percent	Token
329	view	70	0.01	view (63), views (7)
330	flash	70	0.01	
331	he	70	0.01	
332	who	69	0.01	
333	flow	69	0.01	flow (60), flows (8), flowed (1)
No.	Word	Freq.	Percent	Token
334	key	68	0.01	
335	horizon	67	0.01	horizon (10), horizontal (57)
336	another	65	0.01	
337	far	65	0.01	
338	often	65	0.01	
340	scientific	65	0.01	
341	usual	65	0.01	usual (9), usually (56)
342	extreme	64	0.01	
343	poor	63	0.01	
344	river	63	0.01	river (57), rivers (6)
345	although	61	0.01	
346	thus	61	0.01	
347	office	61	0.01	office (59), offices (2)
348	just	59	0.01	
349	without	59	0.01	
350	back	59	0.01	back (57), backed (1), backing (1)
351	necessary	58	0.01	
352	yet	57	0.01	
353	best	56	0.01	
354	half	56	0.01	
	Total	**205,125**	**36.66**	

APPENDIX C

APPENDIX C

The Academic Word List (AWL) Found in the Weather Forecast Corpus (Token)

No.	Lemma	Freq.	Percent	Token
1	area	2,846	0.51	area (1,723), areas (1,123)
2	minimum	1,357	0.24	
3	maximum	1301	0.23	
4	period	938	0.17	period (863), periods (75)
5	data	725	0.13	
6	widespread	562	0.10	
7	issue	516	0.09	issue (32), issues (41), issued (443),
8	technology	469	0.08	technology (450), technologies (19)
9	distribution	461	0.08	distribution (447), distributions (14)
10	ministry	407	0.07	
11	normal	354	0.06	
12	route	330	0.06	route (324), routes (6)
13	index	320	0.06	
14	region	316	0.06	region (161), regions (72), regional (83)
15	proceed	303	0.05	proceed (290), proceeds (30), proceeded (7), proceedings (3)
16	error	286	0.05	error (153), errors (133)
17	research	241	0.04	
18	predict	219	0.04	predict (28) predicts (7), prediction (148) predictions (36)
19	require	211	0.04	require (23), requires (18), required (83), requiring (2), requirement (6), requirements (79)
20	so	177	0.03	
21	available	175	0.03	
22	occur	175	0.03	occur (144), occurs (31)
23	analyze	172	0.03	analyse (1), analyze (3), analyses (69), reanalysis (99)
24	environment	167	0.03	environment (58), environments (6), environmental (103)
25	portion	147	0.03	portion (135), portions (12)
26	method	142	0.03	method (73), methods (69)
27	percent	141	0.03	
28	layer	129	0.02	layer (89), layers (40)

No.	Lemma	Freq.	Percent	Token
29	process	126	0.02	process (65), processes (38), processed (4), processing (19)
30	affect	125	0.02	affect (56), affected (69)
31	define	118	0.02	define (10), defines (3), defined (62), defining (8), definition (16), definitions (19)
32	valid	114	0.02	
33	fee	111	0.02	fee (52), fees (59)
34	approach	106	0.02	approach (73), approaches (23), approached (5), approaching (5)
35	chapter	104	0.02	chapter (95), chapters (9)
36	section	103	0.02	section (83), sections (20)
37	previous	102	0.02	
38	accuracy	101	0.02	
39	source	99	0.02	source (30) sources (69)
40	abstract	95	0.02	abstract (89), abstracts (6)
41	accumulate	93	0.02	accumulate (1), accumulated (12), accumulating (2), accumulation (8), accumulations (1), accumulative (69)
42	energy	90	0.02	
43	project	87	0.02	project (61), projects (26)
44	site	87	0.02	site (59), sites (28)
45	version	86	0.02	version (77), versions (9)
46	intense	81	0.01	
47	significant	77	0.01	
48	specific	69	0.01	
49	sum	61	0.01	
50	potential	59	0.01	
51	constant	58	0.01	
	Total	**15,739**	**2.82**	

APPENDIX D

APPENDIX D

The Technical Word List Found in the Weather Forecast Corpus (Token)

No.	Word	Freq.	Percent	Token
1	forecast	1,746	0.31	forecast (1,746), forecasts (230), forecasted (1), forecasting (106)
2	cloudy	1,605	0.29	
3	knot	1,115	0.20	knot (2), knots (1,113)
4	precipitation	849	0.15	precipitation (847), precipitations (2)
5	meteorology	788	0.14	meteorology (75), meteorological (657) meteor (56)
6	rainfall	623	0.11	rainfall (576), rainfalls (47)
7	climate	591	0.11	climate (534), climatological (57)
8	fog	534	0.10	fog (533), foggy (1)
9	monsoon	503	0.09	monsoon (495), monsoons (7), monsoonal (1)
10	gust	487	0.09	gust (290), gusts (37), gusty (160)
11	atmosphere	381	0.07	atmosphere (147), atmospheric (234)
12	barometer	369	0.07	barometer (224), barometers (145)
13	thunderstorm	284	0.05	thunderstorm (163), thunderstorms (121)
14	humidity	245	0.04	humidity (224), humid (21)
15	hurricane	239	0.04	hurricane (167), hurricanes (66), hurricane's (6)
16	variable	232	0.04	variable (171), variables (61)
17	radar	222	0.04	radar (188), radars (34)
18	satellite	221	0.04	satellite (177), satellites (44)
19	thermometer	217	0.04	thermometer (119), thermometers (98)
20	ice	209	0.04	
21	emission	201	0.04	emission (75), emissions (126)
22	outlook	149	0.03	
23	trough	148	0.03	
24	radiation	146	0.03	radiation (145), radiations (1)
25	ozone	146	0.03	
26	assimilation	131	0.02	assimilation (130), assimilations (1)
27	visibility	123	0.02	visibility (110), visibilities (13)
28	sunset	116	0.02	
29	sunrise	112	0.02	

No.	Word	Freq.	Percent	Token
30	calibration	110	0.02	calibration (105), calibrations (5)
31	haze	108	0.02	haze (82), hazy (26)
32	azimuth	105	0.02	
33	supercell	97	0.02	supercell (62), supercells (35)
34	hail	93	0.02	hail (70), hails (23)
35	velocity	90	0.02	velocity (67), velocities (23)
36	elevation	86	0.02	elevation (86), elevations (3)
37	dataset	86	0.02	dataset (30), datasets (56)
39	moisture	86	0.02	
40	convection	82	0.01	
41	vortex	81	0.01	
42	depression	79	0.01	
43	lightning	75	0.01	depression (74), depressions (5)
44	mist	71	0.01	
45	ridge	71	0.01	ridge (70), ridges (1)
46	dewpoint	71	0.01	dewpoint (67), dewpoints (4)
47	exposure	67	0.01	exposure (65), exposures (2)
48	typhoon	62	0.01	typhoon (58), typhoons (4)
	Total	**14,252**	**2.47**	

APPENDIX E

APPENDIX E

List of Abbreviation Found the Weather Forecast Corpus

Rank	Word	Freq.	Percentage	Meaning
6	°C	4,244	0.76	Celsius
11	km	3,119	0.56	Kilometer
14	hr.	2,765	0.50	Hour
46	Precip/precip	1,239	0.22	Precipitation
47	°F	1,212	0.22	Fahrenheit
63	P.M./p.m.	849	0.15	Post Meridiem
67	AZ/Az	790	0.14	Azimuth
70	A.M./a.m.	751	0.14	Ante Meridiem
10	m	727	0.13	Meter/Metre
75	UTC	727	0.13	Universal Time Coordinated
81	hm/h.m.	687	0.12	Mounting Height
86	No/no.	648	0.12	Number
87	max	637	0.11	Maximum
107	.th	533	0.10	Thailand
114	S	482	0.09	Is, Has, Was
115	min	481	0.09	Minimum
125	H	461	0.08	Height
131	SA	430	0.08	Saturday
132	N	422	0.08	North
138	Fr	412	0.07	Friday
142	Mo	401	0.07	Monday
143	Su	400	0.07	Sunday
90	We	400	0.07	Wednesday
146	Tu	399	0.07	Tuesday
151	LST	384	0.07	Lunar Standard Time
166	NNE	346	0.06	North NorthEast (22.5°)
173	mm	334	0.06	Millimeter
175	CFSR	329	0.06	Contract Funds Status Report
176	E	328	0.06	East =90° (300), Evaporation (1), (e)(27)

Rank	Word	Freq.	Percentage	Meaning
178	NE	325	0.06	NorthEast (45°)
226	Jul	241	0.04	July
234	J	233	0.04	J.(222), (j)(5)
242	Jun	224	0.04	June
259	HPA/hPa	207	0.04	Hectopascals
266	et al.	200	0.04	And others (Et alii)
273	Fig/fig	193	0.03	Figure
274	h.t.	193	0.03	High Tension, High Temperature
281	B	185	0.03	Category B (99), (b) (86)
314	WWW./www.	169	0.03	World Wide Web
317	Lat/lat	165	0.03	Latitude
331	NOAA	159	0.03	National Oceanic Atmospheric Administration
336	WMO	158	0.03	World Meteorological Organization
345	't	154	0.03	not
348	'd	152	0.03	would, had,
354	avg	149	0.03	Average
357	KPH	148	0.03	Kilometers Per Hour
218	Long/long	143	0.03	Longitude
377	NCEP	141	0.03	National Centres for Environmental Prediction
385	K	136	0.02	Kelvin
386	tel	136	0.02	Telephone
388	cm	134	0.02	Centimeter
360	.go.	134	0.02	Government
407	R	119	0.02	Person's Name
30	B.E.	116	0.02	Buddhist Era
430	U.S.	100	0.02	United State
504	TMD	98	0.02	Thai Meteorological Department
507	W	97	0.02	West (270)
518	E.G.	93	0.02	For example
543	A.D.	88	0.02	the Year of our Lord
7	A	87	0.02	(a)
562	Aug	83	0.01	August
584	P	81	0.01	Person's Name
594	TIGGE	79	0.01	THORPEX Interactive Grand Global Ensemble

Rank	Word	Freq.	Percentage	Meaning
616	fax	75	0.01	Photo Facsimile Transmission
632	X	74	0.01	X-ray, X-band, X-Pol
645	L	72	0.01	Person's Name
664	INDO	70	0.01	Indo-China
693	MWR	67	0.01	Mean Width Ratio
699	ENE	66	0.01	East NorthEast (67.5°)
723	NASA	64	0.01	National Aeronautics and Space Agency
745	Jan	61	0.01	January
757	SOC	60	0.01	Society
759	CFS	59	0.01	Climate Forecast System
760	ECMWF	59	0.01	European Centre for Medium-Range Weather Forecasts
	Total	**30,784**	**5.51**	

Appendix F

Appendix F

List of Types of Abbreviations Found in the Study of the Weather Forecast Corpus

1. Clippings

No.	Word	Meaning	Freq.	%
1.	Precip/precip	Precipitation	1,239	0.22
2.	AZ/Az	Azimuth	790	0.14
3.	max	Maximum	637	0.11
4.	We	Wednesday	619	0.11
5.	.th	Thailand	533	0.10
6.	min	Minimum or Minute	481	0.09
7.	Sa	Saturday	430	0.08
8.	Fr	Friday	412	0.07
9.	Mo	Monday	401	0.07
10.	Su	Sunday	400	0.07
11.	Tu	Tuesday	399	0.07
12.	Long/long	Longitude	250	0.04
13.	Jun	June	224	0.03
14.	Fig/fig	Figure	193	0.03
15.	Lat/lat	Latitude	165	0.03
16.	.go	Government	147	0.03
17.	tel	Telephone	136	0.02
18.	Aug	August	83	0.01
19.	fax	Photo Facsimile Transmission	75	0.01
20.	Indo	Indo China	70	0.01
21.	Jan	January	61	0.01
22.	.soc	Society	60	0.01
		Total	**7,805**	**1.36**

2. Initialisms

No.	Word	Meaning	Freq.	%
1.	°C/°c	Celsius	4,244	0.76
2.	km (s)	Kilometer (s)	3,119	0.56
3.	B.E.	Buddhist Era	1,734	0.31
4.	°F	Fahrenheit	1,212	0.22
5.	m (s)	Meter/Metre (s)	727	0.13
6.	UTC	Universal Time Coordinated	727	0.13
7.	hm/h.m.	Mounting Height	687	0.12
8.	' s	Is, Has, Was	482	0.09
9.	H	Height	461	0.08
10.	N	North (0/360°)	422	0.08
11.	LST	Lunar Standard Time	384	0.07
12.	NNE	North NorthEast (22.5°)	346	0.06
13.	mm	Millimeter	332	0.06
14.	CFSR	Contract Funds Status Report	329	0.06
15.	E	East (90°)	328	0.06
16.	NE	NorthEast (45°)	325	0.06
17.	J	Person's Name	232	0.04
18.	HPA	Hectopascals	207	0.04
19.	h.t.	High Tension, High Temperature	193	0.03
20.	B	Type B, (b) / (B)	185	0.03
21.	www.	World Wide Web	169	0.03
22.	WMO	World Meteorological Organization	158	0.03
23.	't	not	154	0.03
24.	'd	had, would, should	152	0.03
25.	KPH	Kilometre Per Hour	148	0.03
26.	K	Kelvin	136	0.02
27.	cm	Centimeter	134	0.02
28.	U.S.	The United States	128	0.02
29.	R	Person'Name	119	0.02
30.	TMD	Thai Meteorological Department	98	0.02
31.	W	West (270°)	97	0.02
32.	P	Person's Name	81	0.01

2. Initialisms (Continued)

No.	Word	Meaning	Freq.	%
33.	X	X-Ray	74	0.01
34.	L	Person's Name	72	0.01
35.	WMO	World Meteorological Organization	69	0.01
36.	MWR	Mean Width Ratio	67	0.01
37.	ENE	East NorthEast (67.5°)	66	0.01
38.	CFS	Climate Forecast System	59	0.01
39.	ECMWF	European Centre for Medium-Range Weather Forecasts	59	0.01
		Total	**18,716**	**3.34**

2. Contractions

No.	Word	Meaning	Freq.	%
1.	hr. (s)	Hour (s)	2,765	0.50
2.	No/no.	Number	648	0.12
3.	et al.	And others/ And elsewhere	200	0.04
5.	avg	Average	149	0.03
		Total	**4,005**	**0.69**

3. Acronyms

No.	Word	Meaning	Freq.	%
1.	NOAA	National Oceanic Atmosheric Administration	159	0.03
2.	NCEP	National Centres for Environmental Prediction	141	0.03
3.	TIGGE	THORPEX (The Observing System Research Predictability Experiment) Interactive Grand Global Ensemble	79 64	0.01 0.01
4.	NASA	National Aeronautics and Space Agency		
		Total	**443**	**0.08**

4. Substitutions

No.	Word	Meaning	Freq	%
1.	P.M./p.m.	Post Meridiem/After Noon	849	0.15
2.	A.M/a.m.	Ante Meridiem/ Before Noon	751	0.14
3.	e.g.	For Example	93	0.02
4.	A.D.	the Year of our Lord	88	0.02
		Total	**1,781**	**0.33**

5. Symbols

No.	Word	Meaning	Freq	%
1.		Thundershowers	720	0.13
2.		Rain	690	0.12
3.		Wave	610	0.11
4.		Temperature	540	0.10
5.		Partly Cloudy	440	0.08
6.		Partly Cloudy	84	0.02
7.		Rainy	210	0.04
8.		Sun	180	0.03
9.		Sunny	180	0.03
10.		Thundershower	156	0.03
11.		Cloudy	130	0.02
12.		Thunderstorm	114	0.02
13.		Windy	112	0.02
14.		Cold	102	0.02
16.		Scattered Rain	96	0.02
17.		Moon	90	0.02

6. **Symbols** (Continued)

No.	Word	Meaning	Freq	%
18.		partly cloudy	84	0.02
19.		wave 1-2 metres	78	0.01
20.		Widely Scattered Rain	72	0.01
21.		wave higher 2 metres	68	0.01
22.		Isolated Rain	63	0.01
23.		cool	57	0.01
24.		cold	56	0.01
		Total	**1,458**	**0.26**

AUTHOR BIOGRAPHY

Name: Miss Siwarak Paengpho

Date of Birth: 19 May 1984

Place of Birth: Nakhonratchasima, Thailand

Education

2002-2003	Certificate in Meteorological Officer from Thai Meteotorological Institue
2002-2004	Bachelor's degree in Political Science, majoring in Public Administrations, minoring in Law from Ramkhumhaeng University
2006-2007	Certificate in English for careers (teaching) from Sukhothai Thammatirat Opened University
2008-2010	Master's degree of Art in Applied Linguistics-English for science and technology from King Mongkut's Institute of Technology Ladkrabang

Work Experience:

2004-2010	Meteorological officer in the Weather Watch and Warning Bureau at Chonburi (Koh Sichang) Meteorological Station.
2010 (November)-present	Meteorological officer in the Northeastern Meteorological Center at Nakhonratchasima (Pak Chong) Agrometeorological Station.