DEVELOPING A MEDICAL ACADEMIC WORD LIST IN ENGLISH: A COMPARATIVE ANALYSIS OF VOCABULARY ACROSS MEDICAL SUB-CORPORA

ANIS ASHRAF ZADEH

FACULTY OF LANGUAGES AND LINGUISTICS

UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

DEVELOPING A MEDICAL ACADEMIC WORD LIST IN ENGLISH: A COMPARATIVE ANALYSIS OF VOCABULARY ACROSS MEDICAL SUB-CORPORA

ANIS ASHRAF ZADEH

THESIS SUBMITTED IN FULFILMENTOF THE REQUIREMENTSFOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Anis Ashraf Zadeh

Registration/Matric No: THA110010

Name of Degree: Doctor of philosophy

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Developing a medical academic word list in English: A comparative analysis of vocabulary across medical sub-corpora

Field of Study: Lexicology

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

This corpus-based lexical study presents the most frequently used medical academic vocabulary across a wide range of medical textbooks to fill the research gap in Wang's et al. (2008) study in developing a medical academic word list based on research articles. The Medical Academic Word List of Textbooks (MAWLOT) was compiled from the Medical Academic Corpus of 3.5 million running words in written medical textbooks by examining the range and frequency of words outside the first 2,000 most frequently occurring words of English (GSL) (West, 1953) and the Academic Word List (AWL) (Coxhead, 2000). MAWLOT contains 459 word families which account for 11.27% of the vocabulary of the Medical Academic Corpus of the study. This good coverage of the MAWLOT provides evidence for teachers and learners in English for Academic Purposes (EAP) and English for Medical Purposes (EMP) to recognize which words are valuable to teach and learn. In addition, concordance lists of the most frequent collocations of the words included in the MAWLOT were provided for studying the words in multiple contexts to observe their behaviours with a greater number of word encounters. MAWLOT is the result of a corpus-based lexical study which, according to Coxhead (2000), "create[s] lists, concordances or data regarding the clustering of linguistic items in coherent, purposeful texts" (p. 227) to provide a useful basis for teachers and learners to set goals in their teaching and learning vocabulary in EAP and EMP. According to Bennett (2010), providing the best possible instruction for students is a desire of many instructors, which is achievable by using corpus linguistics in classrooms. One of the major results of corpus-based studies is developing a wordlist in a specific discipline to help students to practice the frequently used words of the discipline in everyday classrooms (Coxhead, 2010). Offering the first medical academic word list based on the textbooks besides their collocation/phraseological patterns, as well as an extra list of technical words in medicine (53 words) are the main results of

the present study to help medical students fulfil their vocabulary needs in reading medical textbooks written in English.

ABSTRAK

Kajian berdasarkan korpus leksikal ini memperkenalkan senarai kosa kata perubatan akademik yang paling kerap ditemui dalam pelbagai buku teks perubatan untuk menutupi jurang dalam kajian Wang et al. (2008) yang menerbitkan senarai kosa kata perubatan akademik berasaskan rencana-rencana ilmiah. Senarai Kata Akademik Perubatan daripada Buku-Buku Teks (MAWLOT) disusun dari Korpus Perubatan Akademik berjumlah 3.5 juta perkataan yang ditemui dalam teks perubatan akademik dengan cara memeriksa julat dan kekerapan perkataan di luar 2,000 perkataan Inggeris yang paling kerap (GSL) (West, 1953) dan Senarai Kata Akademik (AWL) (Coxhead, 2000). MAWLOT mengandungi 459 keluarga perkataan yang mencakupi 11.27% daripada vokabulari Korpus Akademik Perubatan yang digunakan dalam kajian itu. Liputan MAWLOT memberi peluang kepada guru dan pelajar EAP (Bahasa Inggeris untuk Tujuan Akademik) dan EMP (Bahasa Inggeris untuk Tujuan Perubatan) mengenalpasti perkataan-perkataan yang perlu untuk diajar dan dipelajari. Tambahan lagi dengan tujuan menyerlahkan kata-kata yang dimasukkan ke dalam MAWLOT, maka senarai konkordans daripada kolokasi perkataan yang paling kerap disediakan untuk mempelajari kata-kata dalam pelbagai konteks untuk melihat peranannya dengan jumlah ungkapan kata yang lebih besar. MAWLOT adalah hasil kajian leksikal berdasarkan korpus yang menurut Coxhead (2000) "mewujudkan senarai, konkordans atau data yang mengelompokkan item linguistik dalam teks yang koheren dan bermatlamat" (hlm. 227) bagi menyediakan bahan yang berguna untuk guru dan pelajar demi menentukan tujuan dalam pengajaran dan pembelajaran kosakata dalam EAP dan EMP. Menurut Bennett (2010), ramai pengajar mahukan pemberian arahan yang terbaik kepada pelajar-pelajar dan ini boleh dicapai dengan menggunakan pendekatan linguistik korpus dalam bilik darjah. Salah satu hasil utama kajian berasaskan korpus adalah penyusunan senarai perkataan dalam bidang khas untuk membantu pelajar

menggunakan kata-kata dengan kekerapan tinggi daripada bidang itu dalam bilik darjah setiap hari (Coxhead, 2010). Senarai kosa kata perubatan akademik yang pertama berdasarkan buku-buku teks ini berserta dengan kolokasi kata-kata serta senarai tambahan perkataan teknikal dalam perubatan (53 perkataan) dengan kolokasinya adalah hasil utama kajian ini yang akan membantu pelajar perubatan untuk memahami dengan lebih baik teks-teks perubatan bahasa Inggeris.

ACKNOWLEDGEMENTS

In memory of my brother "Omid" (hope), this research is dedicated to:

My great parents, Professor Dr. Reza Ashraf Zadeh and Tahereh Sayfaddini, who always love and support me with no limitations.

This thesis could not have been completed without the guidance, advice, support and encouragement from many people. My greatest of debts is to my dear supervisors, Associate Professor Dr. Victor A. Pogadaev and Associate Professor Dr. Su'ad Awab for safely guiding in writing this thesis and for encouraging me to keep going.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AWL	:	Academic Word List (Coxhead, 2000)
BNC	:	British National Corpus
CLT	:	The software package of Compleat Lexical Tutor
EAP	:	English for Academic Purposes
EGAP	:	English for General Academic Purposes
ELL	:	English Language Learner
EMP	:	English for Medical Purposes
EOP	:	English for Occupational Purposes
ESP	:	English for Specific Purposes
ESAP	:	English for Specific Academic Purposes
GP	:	General Physician: A medical student in his or her first stage of studying medicine, which is about 7 years in Iran.
GSL	:	General Service List of English Words
ICLE	:	The International Corpus of Learner English
IMGs	:	International Medical Graduates
IMRD	:	Introduction-Method-Result- Discussion
L1	:	The first language
L2	:	The second language
MAWL	:	The Medical Academic Word List of research articles (Wang et al., 2008)
MAWLOT	:	The Medical Academic Word List of Textbooks
MAWLOT+	:	The Medical Academic Word List of Textbooks besides the medical technical words
NNES	:	Nonnative English Speakers

SLVA	:	Second-Language Vocabulary Acquisition
SST	:	The Standard Speaking Test Corpus

CHAPTER 1: INTRODUCTION

1.1 Overview

Researchers have previously produced lists of academic vocabulary words focusing on single disciplines. For example, in Computer Science a list of vocabulary words was developed by Lam (2001) to find problems of vocabulary learning which computer science students encountered while reading academic texts. In engineering, Mudraya (2006) compiled an academic word list of 1,200 word families based on textbooks in 13 engineering subjects. This is an important word list containing about 2 million running words. Similar research has also been carried out by Vongpumivitch, Huang and Chang (2009) in applied linguistics.

In the area of medicine, the Medical Academic Word List (MAWL) was established by Wang et al. in 2008. It was the second effort, after an earlier lexical study conducted by Chen and Ge (2007) to produce an academic word list in the field of medicine. Chen and Ge conducted a lexical study on word frequencies and text coverage (range) of the 570 word families of the Academic Word List (AWL). By developing a corpus of medical research articles with 190,425 running words (tokens) taken from 50 medical research articles written in English, they investigated the text coverage of the Academic Word List (AWL) (Coxhead, 2008) throughout the medical research articles. As a result, Chen and Ge found that the AWL was not able to demonstrate the most frequently used medical academic vocabulary words in medical research articles. Therefore, they called for efforts to develop a medical academic word list for use by medical students for reading written texts. This eventually led Wang et al. to establish the Medical Academic Word List (MAWL) in 2008. The list was retrieved from a corpus of 1,093,011 running words in medical research articles from online resources, and contained 632 word families covering 12.24% of the corpus. The list of Wang et al. (2008) is extremely useful in designing language courses for Medical English; however,

there is still a research gap in the area of medicine (EMP) in developing an academic word list based on textbooks, as Wang et al. (2008) mentioned. Filling this research gap is the main goal of the present study by developing a medical academic word list from textbooks.

1.2 Statement of the problem

As Wang et al. (2008) admit, their research is only "preliminary" (p. 452) in nature, and there is still a research gap in developing a medical academic word list based on textbooks. The principle research goal of this study is to fill this research gap. In fact, Wang et al. (2008) go on to say: "If possible, the MAWL needs to be rechecked in larger corpora or in other genres of medicine, such as medical textbooks or spoken medical academic English" (p. 452). According to Wang et al. (2008), few detailed studies have been done exclusively on medical academic vocabulary used in the discipline of medicine.

In addition, as Lei and Liu (2016) mention "it is important to note that Wang et al. (2008) did not include any medical English textbook data in their study" (p. 46), while the main users of medical vocabulary lists are medical students who "have to read medical textbooks in their study" (p. 46). Lei and Liu believe that to get a good understanding of the typical lexical characteristics of medical discourse, attention needs to be paid to medicine as an academic discipline with its own discourse community, with its own genres and conventions. They also state that, although research articles are an important academic genre, there is still a need for the study of medical textbooks. After all, English-language textbooks are used in many universities and medical schools across the world, not only by lecturers, but also by students and even by language instructors to determine what vocabulary to teach in their English for Medical Purposes (EMP) courses.

Furthermore, it is clear that developing an English-language medical academic word list based on medical textbooks will have many benefits. For example, it will enable researchers, course designers and teachers involved in the area of English for Academic Purposes (EAP) to provide medical students with vocabulary lists (e.g. the AWL (Coxhead, 2000), and hence, more effective syllabi and course materials.

1.3 Objectives of the study

This PhD research project seeks to attain two objectives. The first objective is to develop a Medical Academic Word List of Textbooks (MAWLOT). Reviewing studies in developing academic word lists in specific genres reveals a lack of achievement in developing an academic word list of textbooks in medicine (Wang et al., 2008). It is crucial that these medical textbooks are not chosen at random from the publishers, but are taken instead from the titles used in and suggested by authoritative English-speaking medical universities, which are considered the main references in medical universities in Iran. Selecting the textbooks from authoritative English-speaking medical universities enables me to compile a representative corpus to develop a medical academic corpus of textbooks. Representativeness is an important criterion to consider in developing any corpus (Biber, 1993) to allow findings to be generalized to a particular variety of language.

In addition, identifying the most frequent collocation/phraseological patterns of the words included in the MAWLOT will be done to observe the words' behaviours in their contexts. Also, as Nation and Waring (1997) point out, vocabulary learning needs to have clear and sensible aims. So, the underlying assumption is that a MAWLOT of the most frequently used vocabulary will help define those aims, and that it will facilitate the process of learning for students and teaching for instructors. The lexical analysis of our corpus will also allow us to 'recheck' Wang's et al. (2008) MAWL and, where

relevant, similarities and differences will be reported and explained to find out how a medical academic word list of textbooks, such as the present study, can be applied in EAP/EMP to increase the students' vocabulary and reading comprehension.

The second objective of this study is to investigate, in more depth, the similarities and differences in frequencies between the lexical items that characterize the various sub-sections of our corpus of medical textbooks. These sub-sections will consist of more specific medical disciplines; i.e. the major fields in any medical college or university degree program. This comparative component will enable me to establish the range of certain lexical items across various medical specialties and practices. Like frequency, range is a critical factor in deciding what words to target in EAP (Sinclair, 1991).

Identifying the major fields of medicine and recognising the latest reference (s) for the various major fields, are considered in the Methodology chapter.

1.4 Significance of the study

The significance of this study is twofold. Firstly, as mentioned above, there is a real research gap in the lexical study of medical English, meaning that this study will advance our *descriptive knowledge* of this variety. Despite an abundance of vocabulary research in disciplines like business, management and engineering there are comparatively few studies other than those of Chen and Ge (2007) and Wang et al. (2008) (recently the study of Lei and Liu (2016) has also been done in the area of medicine), exclusively interested in medical academic vocabulary. Furthermore, most of these studies are based on research articles only, and most treat medical vocabulary as a unified phenomenon. In addition, few or no distinctions are made between such disciplines as, for example, internal medicine and surgery or pediatrics. A descriptive word list containing frequent words occurring in a wide range of medical textbooks will

help to identify academic vocabulary for instruction, i.e. which words are valuable enough to teach, and will help students to recognize which words are valuable enough to learn.

Therefore, this study will greatly contribute to the research fields of EAP and in particular EMP (English for Medical Purposes). First-year medical students, like other students that are new to a discipline, have to become proficient in a core of high-frequency academic words (Hyland & Tse, 2007) to attend lectures and tutorials in L2, read and study textbooks in L2, and talk to professors in L2. The relevance of this core vocabulary knowledge to academic performance and to their competence in many academic genres (e.g. case reports and research presentations) cannot be underestimated, because these activities are directly related to the main tasks of students through their studies in universities. It is evident that a detailed corpus-based lexical study is a reliable and accurate method to determine the most frequent medical academic words in medical textbooks, not only in general but also in terms of medical specialties and practices. Another area that relates to our study has to do with the notion of collocation. Schmitt (2004) and others before him point out that a learner's L2 fluency depends to a large extent on his or her understanding and usage of collocations; i.e. multi-word expressions. For example, it is not enough to know the noun *muscle* or its derived adjective *muscular*. A medical student also needs to know which one to use with the word *spasm* to form a high-frequency, multi-word phrase. Corpus analysis will then be useful in deciding whether *muscle spasm* or *muscular spasm* is more frequently used.

Meanwhile, this study seeks to develop a Medical Academic Word List of Textbooks (MAWLOT) by employing academic software tools, i.e. the Compleat Lexical Tutor (CLT) version 8, and the AntConc software version 3.4.4), which have

been designed exclusively for lexicon studies in all genres. The focus of CLT is on academic vocabulary; therefore, the software program and its many features have great potential for those working with English for Academic Purposes (EAP) (Cobb & Horst, 2004). CLT has been used in the study of Farjami (2014). In addition, the AntConc software has been extensively used in collocations studies by researchers. For example, in the study of Noguchi (2004) it has been described as extremely effective in a classroom context.

Secondly, the significance of this study will not only be in terms of sampling authentic and representative lexical data, describing these data, and explaining the main lexical trends across various medical subjects where possible. This study also has *practical* relevance in the areas of EAP and EMP course design and teaching. The importance of frequency-based word lists, as the lists that contain very common words, rests on the belief that the lists inform learners which words are valuable to learn (Cobb & Horst, 2004). Similar observations have also been made by Nation and Waring (1997). Summing up their main argument, frequency information provides a reasonable basis for making sure that a learner gets the best return for his or her learning of vocabulary. It follows that providing a vocabulary frequency list can have an important role to play in setting learning goals and curriculum design in all academic genres and disciplines, not only in medicine.

Obviously, learners in an academic area do not need to know, let alone memorize, long lists of monolingual decontextualised words. According to Hyland and Tse (2007), vocabulary taught in schools is the vocabulary critical to understanding the concepts of the content; this vocabulary is thus defined as academic vocabulary. Identification of academic vocabulary requires knowing the degree of the importance of the terms. Some terms are useful, but are not critical. Some terms are not critical, but are interesting. The

same statements for lexical phrases and collocations, beyond the single words, can be made, too.

Rather, referring to specific lists for course designers when they consider the vocabulary component of an EAP or EMP language course could be important. Also, teachers need to have reference lists to judge whether or not a particular word deserves explicit attention in class. The word lists are thus helpful in creating writing course materials and in shaping classroom input of the kinds discussed, for example, in Milton (2009, Chapter 9). With the above insights in mind, as a starting point, the present study aims to explore answers to the following research question:

1.5 Research questions

The purpose of this study is to develop and evaluate a medical academic word list based on textbooks (MAWLOT). Three questions frame the description of the MAWLOT:

1. What are the MAWLOT's lexical items in terms of frequency and range in the corpus of medical textbooks?

2. Do the MAWLOT's lexical items reveal particular patterns in the corpus of medical textbooks?

3. Do the MAWLOT's lexical items occur with different frequencies in the medical textbooks?

The evaluation of the MAWLOT also considers the following three questions:

4. What proportion of the lexical items in the medical academic corpus appear in the MAWLOT?

5. Do the lexical items identified appear often in a different medical text collection?

6. How does the MAWLOT compare with the AWL (Coxhead, 2000) and the MAWL (Wang et al., 2008) in terms of tokens, types, number of words included in the wordlists, and the coverage of the word lists of the vocabulary in their corpora?

1.6 Limitations of the study

This study has three limitations.

1. This study is based on a single medical discourse genre, i.e. medical textbooks, and does not consider other disciplines like spoken medical academic English, case reports, etc.

2. The focus of the study is on developing a medical academic word list based on the textbooks for general physicians (GPs), especially in Iran, and does not consider the textbooks for specialists like internal medicine specialists, gynecologists, pediatricians, psychiatrists, and surgeons. This limitation enables a wide range of medical students, i.e. general physicians, to benefit from the word list in their reading comprehension of medical textbooks.

3. Similar to researchers who are interested in learning words in instructional contexts with more focus on word frequency and their corpora collocations, and less concentration on demonstration and processing of richer semantic representations (Nation, 2001; Cobb & Horst, 2004), the focus of this study is on describing the frequencies of words and the frequent collocations of the words across a variety of subdisciplines of medicine.

1.7 Definitions of terms

This section presents definitions of terms and concepts that have a key role in the current study are presented. The aim of this short section is to clarify the scope of the present study and the limits within which the findings of the investigation should be interpreted.

Academic Corpus: The corpus of 3,500,000 running words in 4 subdisciplines of arts, commerce, law, and science to develop the Academic Word List (AWL) by Coxhead (2000).

Academic vocabulary: Vocabulary frequently found in different written academic text types such as textbooks, handouts, manuals, articles and research abstracts. The term 'academic vocabulary' is used in the present study to refer to both the general and subject-specific vocabulary used in an academic context.

Academic Word List (AWL): A list of 570 words developed from a corpus of 3,500,000 running words (Academic Corpus) in four subdisciplines of arts, commerce, law, and science in 2000 by Coxhead.

The AntConc: The software program version 3.4.4, which is available at the URL address http://www.laurenceanthony.net/software/antconc/.

Corpus: Generally, corpus (plural: corpora) refers to a collection of texts that plays a role as the bodies of evidence in linguistic analysis (Thompson, 2006). This term is also defined as "a body of written text or transcribed speech which can serve as a basis for linguistic analysis or description" (Kennedy, 2014, p. 1), and a "large principled collection of naturally occurring examples of language stored electronically" (Bennett, 2010, p. 2).

The Compleat Lexical Tutor: The Compleat Lexical Tutor (CLT) version 8 is a software package that is used extensively to perform lexical text analysis to measure the proportions of low-frequency and high-frequency vocabulary used by speakers of varying ability. The software program is available at the URL address: *http://www.lextutor.ca.*

EAP: English for Academic Purposes (EAP) is the major branch of English for Specific Purposes (ESP), which has involved practical experiences in both English-speaking countries and many other countries in which English is the language of research in the

sciences (Hutchinson & Waters, 1987). In this study EAP is used as an umbrella term for both English for General Academic Purposes (EAGP) and/or English for Specific Academic Purposes (ESAP).

Genre: This study follows the definition of Hyland (2004) in defining 'genre' as "a term for grouping texts together, representing how writers typically use language to respond to recurring situations" (p. 4).

General physician (GP): A medical student in his or her first stage of studying medicine, which is about 7 years in Iran. During this stage of study, medical students will be familiarized with all medical subject areas based on their universities' curriculum in general rather than specific terms.

Medical Academic Corpus: The corpus of the current study, which contains 3,500,000 running words of 5 subdisciplines of medicine considered in the present study, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery to develop the Medical Academic Word List of Textbooks.

MAWL: The Medical Academic Word List developed by Wang et al. (2008) from a corpus of 1,093,011 running words of online medical research articles (RAs). The MAWL contains 623 word families, which accounts for 12.24% of the tokens (running words) in the corpus of medical RAs.

MAWLOT: The Medical Academic Word List of Textbooks of the present study developed from a corpus of 3.5 million running words of medical textbooks. The word list contains 459 words that cover of 11.27% of the vocabulary of the Medical Academic Corpus of the study.

MAWLOT+: The Medical Academic Word List of Textbooks (459 words) in addition to a list of medical technical words (53 words) of the present study.

Medical technical vocabulary list: A list of medical-specific words (53 words) of the present study, such as *polynomial* and *cytoplasm* that are used only in medical contexts and texts.

Medical technical words: The words those are particularly unique in medicine. They are the discipline-specific words and have just one meaning in medicine, such as *anemia, carcinoid,* and *hemoglobin.*

Technical vocabulary: The words that are particularly unique and are used only in one discipline like medicine, math, or science. These words are counted as the discipline-specific words (domain-specific academic words or content-specific words) (Hiebert & Lubliner, 2008). They have also just one meaning, the understanding of which is essential for building a conceptual knowledge in the discipline they are using.

Tokens: This term refers to total numbers of word forms (running words) in a text.

Types: This term refers to each different word (individual words) in a text, irrespective of how many times it occurs (Coxhead, 2000).

Word list: A list which presents the words occurring in a corpus along with the number of times each word appears (Bennett, 2010).

Word family: A stem besides its inflected and derived forms, and all closely related affixed forms plus the most frequent stem, perceived transparency, and regular and productive prefixes and suffixes (Bauer & Nation, 1993).

1.8 Outline of the thesis

This chapter describes the overview of the study concerning the importance of developing a medical academic word list based on textbooks in reading comprehension in medicine, as well as the objectives and limitations of the study.

Chapters II and III review the underlying principles that lend support to the application of developing academic word lists in vocabulary learning. Chapter II starts with reviewing the concepts of English for Academic Purposes (EAP) and the

background studies on developing academic vocabulary lists in EAP/ESP, highlighting the role of academic word lists in enhancing the students' vocabulary knowledge. Next, the classifications of vocabulary with focus on the type of academic vocabulary are described. Chapter 3 focuses on corpus linguistics as the most related theory to the present study while reviewing previous studies in corpus analysis to develop academic word lists. Finally, the stages in developing the corpora and wordlists of the AWL (Coxhead, 2000) and the MAWL (Wang et al., 2008) are reviewed to find out the principles for compiling the Medical Academic Corpus of the present study in developing Medical Academic Word List of Textbooks (MAWLOT). The chapter ends with considering some notions related to corpus linguistics, including phraseology and collocation, to gain a view of words by considering them in their contexts.

Chapter 4 deals with the research methodology of the study, which includes population and samples of the study as well as the pilot study with proposed changes for the main study, the method of data collection and research instruments. The method of data analyses is described at the end of the chapter. Chapter 5 presents the findings and focuses on the two main phases of the study, i.e. the descriptive phase and evaluation phase. Chapter 6 presents the patterns that are associated with particular lexical items in the Medical Academic Corpus of the study. The role of collocations in teaching language besides discussions related to the findings will be provided by the chapter.

In Chapter 7, a summary of the present study is provided prior to discussions of the findings of the two phases of the study, its pedagogical implications, and recommendations for further research.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In this chapter, the literature related to the study is reviewed to show how the current study builds on earlier studies in various research areas in developing academic word lists and also to obtain a theoretical framework for developing a medical academic word list based on textbooks. The review is divided into three main areas: (i) English for Academic Purposes, (ii) academic vocabulary, and (iii) corpus linguistics approach.

Firstly, English for Academic Purposes (EAP) is considered in terms of the importance of EAP courses in the university curricula all over the world. The background studies on developing word lists in EAP are presented to highlight the importance of corpus-based studies in developing word lists, and to consider the usefulness and/or limitations of word lists in EAP learning.

Secondly, academic vocabulary is identified in terms of the roles of different vocabulary types in developing word lists across different disciplines. Finally, the corpus linguistics approach is reviewed in order to obtain a framework compatible with developing a medical academic word list from textbooks.

2.2 English for Academic Purposes

The role of English as the lingua franca of science has increased the value of English for Academic Purposes (EAP) in university curricula all over the world. For several years EAP has been the major branch of English for Specific Purposes (ESP) not only in English-speaking countries, but also in many countries in which English has been the language of research in sciences (Hutchinson & Waters, 1987).

The term ‘English for Academic Purposes’ (EAP) was used for the first time in 1974 (Johns, 1990), and was extensively investigated by Swales in 1990. His investigations

included academic reading and writing with a special focus on written graduate theses and dissertations. Dudley-Evans and St. John (1998) considered four main geographical domains for EAP: (a) the major English speaking countries, like the UK, USA, Canada, New Zealand and Australia, where English is not the first language of large numbers of overseas students; (b) the former colonial territories of Britain, like India, Malaysia and Singapore, where English is the second language and language of higher education; (c) the countries with no historical links to the UK but that use English to access research, like Japan, China, Latin America, and the countries of Western Europe; and (d) the former Soviet-bloc countries that are increasingly taking part in the academic community and global economy.

Hyland and Hamp-Lyons (2002) considered the efforts to study EAP during the past 25 years and concluded that over the last 25 years EAP has experienced fast progress and has become a major force in studying language teaching. On the other hand, Watson Todd (2003) mentioned that the studies on EAP during the last 25 years have focused more on the content of teaching than on methodology. In addition, Hyland and Hamp-Lyons emphasized the role of EAP in showing the barriers of social contexts in language use, and that there is a need to find new ways to gain control over these barriers for EAP learners. However, Watson Todd (2003) mentioned that the main role here has been played by ESP, and EAP is just one of the two major branches of English for Specific Purposes (ESP). The other major branch of ESP is English for Occupational Purposes (EOP), with subdivisions such as English for Pilots, English for Doctors, English for Bank Employees, English for Nurses, etc. The subdivisions of EAP are English for Medicine, English for Mathematics, English for Economics, and English for Biology. EAP practitioners work in academic places such as research institutions, while workers in EOP are located in professional workplaces.

In addition, language researchers believe that EAP is an international phenomenon that has become important in terms of communication, education, and globalization of information exchange (Douglas, 2000; Hutchinson & Waters, 1987). According to McKenny (2006), assisting students to learn how to use their abilities in listening, speaking, writing, and reading in their educational lives and future jobs is the main goal of teaching EAP. For university-level students the knowledge of academic vocabulary is essential to the process of learning (Coxhead, 2000) by means of an EAP course affects academic communication skills and overall academic performances (Hyland & Tse, 2007). Besides, "an active vocabulary teaching program may make students more 'word conscious', thus motivates them to learn more words through context" (Stahl, 1986, p. 667).

Furthermore, the ineffectiveness of incidental learning and difficulties experienced by many university students, besides the pervasiveness of academic vocabulary across disciplines and genres, are all convincing reasons for offering EAP courses, both general and subject-specific. An EAP course itself, as a pedagogical intervention, is a way in which college/university students can acquire an academic vocabulary (Milton, 2009). This course is meant to benefit students who are new to a genre (e.g. review article) or discipline or are not proficient in reading academic texts and have some difficulties in discourse comprehension. According to Milton, textbooks are an extremely effective way to introduce a large quantity of new vocabulary words to learners if certain conditions are met. For example, vocabulary must be introduced at regular intervals, and frequent words, infrequent words and multi-word expressions must be presented in roughly equal amounts.

Regarding the present study, developing a word list by using medical textbooks as the main references to identify and highlight the most frequent words occurring in a

wide range of texts will help instructors and EAP (EMP) teachers to recognize the words worthy of emphasis. Studying these words will facilitate the process of learning and increase the students' lexical knowledge that will be discussed in the next sections of the chapter.

2.3 English for Medical Purposes (EMP)

Some researchers believe that English for Medical Purposes (EMP), like English for Specific Purposes (ESP) and English for Science and Technology (EST), includes a large area of learning, instructions and research paradoxes (Rebek-Nagy et al., 2008). Rebek-Nagy et al. define EMP as a form of the second language of education that obviously focuses on the teaching aspects of medical English to fill job needs such as interviewing and assessing patients, reporting, writing progress notes, and charting. According to Rebek-Nagy et al., in addition to EMP students such as physicians, nurses, and pharmacists, there are also the International Medical Graduates (IMGs) who have been trained outside the English language countries. These IMGs are mostly physicians who work in the fields of pediatrics, psychiatry, internal medicine and family medicine, and they provide a unique opportunity for ESL/ESP professionals to contribute to medicine by improving communication between the IMGs and their patients.

In addition, Frank (2000) mentions that an area like medical vocabulary needs more attention. Frank notes that there should be a bottom-up approach in the teaching of medical vocabulary through teaching prefixes, suffixes, word derivation and roots with an emphasis on different vocabulary needs in different situations. Frank believes that three perspectives have guided research in EMP: (a) medical professionals who are not from English-speaking countries but need to learn English to practice their professions in native-speaking countries, like the UK and USA (Candlin, 1981); (b) medical professionals who desire to publish their research in English medical journals or to

participate in international research and conferences (Allwright, 1984; Salager-Meyer, 1990; Chandler-Burns, 1997), and (c) university students who plan to study medicine in an English-speaking country (Graham & Beardsley, 1986). These perspectives deal with the traditional view of ESP, which concentrates on the communication needs of Non-native English Speakers (NNES) in their work and studies.

Therefore, the main goal of the present study is to help medical students, especially the ESL students fulfil their vocabulary needs of medical texts by recognizing the most frequent words that occur in a wide range of medical textbooks.

2.4 Background studies on developing academic vocabulary lists in EAP/ESP learning

For different academic disciplines, researchers have previously produced lists of academic vocabulary. One of the earliest endeavours was Thorndike, Lorge and their colleagues (1944), who based their work on 18,000,000 running words, which they sifted through manually. Corpus research into academic English began in earnest, as of the 1970s. Campion and Elley (1971) analysed 301,800 words from 19 academic disciplines in both textbooks and journals, and retrieved the 500 most common words and the 3,200 most frequently used words. The following year, the American University Word list was published by Praninskas (1972). He extracted the list from 272,466 words of 10 academic disciplines in 10 university-level textbooks. Lynn (1973) and Ghadessy (1979) listed the words that foreign students had difficulty with in their reading and also examined the annotations written in their university textbooks.

The University Word List (UWL) was drawn up and published by Xue and Nation in 1984, consisting of about 800 words that included high-frequency and wide-range words in academic texts. They combined the earlier word lists (4 word lists) complied by Campion and Elley, Lynn and Ghadessy, Praninskas and Xue and Nation,

respectively, to produce a high-frequency word list for learners in any academic discipline, which could be taught to them directly.

Coxhead's (2000) Academic Word List (AWL) is another well-known academic word list. To develop the AWL, she used the Range and the State-of-the-art computer software to calculate a word's frequency (how often a word occurs) and range (in how many different texts in the corpus it occurs). The list was retrieved from a corpus of about 3.5 million running words selected from different university textbooks and academic journals. The AWL consisted of 570 word families (about 10% of the corpus) in four main areas: natural science, arts, law and commerce. After starting the present study, a new Academic Vocabulary List (AVL) was developed by Gardner and Davies (2013) from a 120-million-word academic corpus. "Since this core academic vocabulary list is new in the field of academia, empirical studies based on the evaluation of this list are not yet available" (Hajiyeva, 2015, p. 137).

Recently, researchers have focused more on a single discipline to develop academic vocabulary lists. An example is a list of 729 most frequent word families in engineering compiled from engineering textbooks by Hsu in 2014. Before the Hsu's list, the Engineering Academic Word List was developed by Mudraya in 2006. Mudraya established the Student Engineering English Corpus (SEEC) with about 2 million running words to compile an academic word list of 1,200 word families, based on textbooks in 13 engineering subjects.

Similar research has also been carried out in the discipline of applied linguistics. Vongpumivitch, Huang and Chang (2009) compared the frequency of words included in the AWL in research articles of applied linguistics journals, namely *Applied Linguistics, The Modern Language Journal, Second Language Research, TESOL Quarterly*, and *Language Learning.* In the field of agriculture, Martinez and her colleagues also used

the AWL in 2009 as their starting point to develop a restricted list of 92 family words in agricultural research articles. Finally, in the discipline of finance, Li and Qian (2010) profiled the presence of the AWL items in the Hong Kong Financial Service Corpus (HKFSC) to find ways to teach the AWL items attested in the corpus efficiently.

As for the subject of the present study, i.e. medicine, Chen and Ge (2007) conducted a lexical study on word frequencies and text coverage (range) of the 570 word families of the AWL in a corpus of medical research articles. They compiled a corpus with 190,425 running words taken from 50 medical research articles written in English. Although their study confirmed the text coverage of the AWL throughout the medical research articles, they called for efforts to develop a medical academic word list, as they found that the AWL was not able to demonstrate the most frequently used medical academic vocabulary in medical research articles. This eventually led to the establishment of the Medical Academic Word List (MAWL) by Wang et al. in 2008. The list was retrieved from a corpus of 1,093,011 running words of medical research articles from online resources. Recently, another study has been done by Lei and Liu (2016). The study focuses on developing a medical word list based on both research articles and textbooks. The resulted list contains 819 lemmas. This study has also been done after starting the present study which focuses just on the medical textbooks. Similar to the study of Gardner and Davies (2013), the empirical studies based on the evaluation of this list have not been done yet.

As a result, what is crystal-clear from all these lexical studies is the interest in working on the vocabulary of one specific academic discipline only. It also seems that the analysis of academic vocabulary has become more and more specialised and focused, and it is to this trend that the current study on medical English hopes to make a valuable contribution. It is obvious that there is a research gap in the area of medicine

(EMP) in developing an academic word list based on textbooks. The only word list in medicine is the medical academic word list (MAWL) established by Wang et al. (2008) based on a corpus of medical research articles (RAs), which contains 632 word families with a coverage of 12.24% of the corpus. This research gap has been echoed by Wang et al. (2008) as they state: "If possible, the MAWL needs to be rechecked in larger corpora or in other genres of medicine, such as medical textbooks or spoken medical academic English" (p. 452). Therefore, the main goal of the present study is to fill the second part of Wang et al.'s research gap, i.e. developing a medical academic word list based on medical textbooks.

2.5 Academic vocabulary

It is generally accepted that lexical knowledge is a necessary factor for successful reading comprehension. To increase students' lexical knowledge in reading medical texts, the exact words must be identified in order to ensure that they are worth being studied. Accordingly, consideration must be taken towards vocabulary types to obtain the principles for selecting words most suitable for inclusion in the Medical Academic Word List of Textbooks (MAWLOT). Evaluating these words will be important, too, which is considered extensively in the Methodology Chapter of the study.

Academic vocabulary has been defined by Farrell (1990, p. 11) as "formal words with a high frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses; words with a high frequency across scientific disciplines". A synonym for 'academic" is "sub-technical" (Laufer & Nation, 1999). Following Laufer and Nation (1995), it is usually distinguished from more technical, subject-specific vocabulary that is part of a discipline or field of enquiry like biomechanics or film theory, and also from general service vocabulary (very high-frequency general English). Academic vocabulary, as many researchers believe, is a

practical tool for students of college/university-level academic subjects like medicine, media studies or musicology. More generally, since academic vocabulary is the language of learning (Hyland & Tse, 2007), teaching it explicitly by means of an EAP course can be expected to facilitate the learning process and to affect academic communication skills and overall academic performance. Obviously, in EAP as in any other type of formal teaching, students are not likely to pick up academic vocabulary automatically or effortlessly. So-called incidental vocabulary learning is very ineffective, as Martin (2009) has recently shown. Without explicit, well-structured and focused instruction, students will not learn what, for example, a compound noun like *stem cell* means or which verbs collocate with *dose*, the 12th item in Wang's et al. MAWL (2008).

Moreover, it is generally accepted that the use of academic vocabulary is problematic for many language learners (Li & Pemberton, 1994; Thurston & Candlin, 1998; Cohen et al., 1988; Leeke & Shaw, 2000; Coxhead & Nation, 2001; Chung & Nation, 2003; Cobb & Horst, 2004). Understanding the teacher's explanations, discussing them, reading to get more information in the area, and writing about what they have learned are the students' needs in an academic context. Understanding the concepts of the academic words as used throughout the course also presupposes breadth and depth of technical vocabulary knowledge. Since comprehension directly improves with expansion of vocabulary (Stahl & Fairbanks, 1986), it is important for students to have a deep understanding of the technical content vocabulary that is specific to the subject to master the main and expected vocabulary concepts throughout the course. For example, a business English word like *turnover*, which has four or five meanings, would throw students off track and they could not understand the meaning of a sentence or text in which it appears. These issues not only affect L2 students who are studying EAP

courses in English-speaking countries, but also those who have to attend lectures, tutorials in L2, read and study textbooks in L2 and talk to professors in L2.

Regarding the present study in developing the Medical Academic Word List of Textbooks (MAWLOT), the definition of Laufer and Nation (1999) which applies the 'academic' and 'sub-technical' words as synonym terms, however distinguishable from technical or subject-specific vocabulary that is part of a discipline or a field of inquiry, is used. In addition, identifying the academic words which occur frequently across a wide range of medical textbooks besides their frequent lexical phrases and collocations may help L2 learners to increase their vocabulary knowledge awareness.

2.5.1 Types of vocabulary

Apart from the different definitions of academic vocabulary and considering academic vocabulary as a main part of the study in developing word lists, it is obvious that some vocabulary is used more frequently than other words in different academic contexts, or they are found more frequently in textbooks than in articles. Also, some vocabulary is used more often in medical textbooks than in science or engineering ones, for example. Previous studies show that those words that occur frequently in students' academic texts and are relevant to their needs are identified and classified for pedagogical use (Nation, 2000). Therefore, to obtain the principles to select the most suitable words for inclusion in the Medical Academic Word List of Textbooks (MAWLOT) of the present study, consideration must be taken towards the four main types of vocabulary: high-frequency words, low-frequency words, academic words and technical words (Nation, 2001; Coxhead & Nation, 2001; Dudley-Evans & St. John, 1998; Nation & Waring, 1997). Selecting the words to teach in EAP/EMP courses is important to developing any academic word list. Increasing medical students' lexical

knowledge by paying more attention to the selected words included in MAWLOT is the main goal of the present study.

2.5.1.1 High frequency words

High-frequency words are a group of 2,000 word families (West, 1953) that is considered to be the essential basis for all use of language. A very large proportion of the running words in spoken and written texts, like *question*, *work*, *low*, *right*, *metal*, *paper*, *usually*, etc. belongs to this group. Prepositions, conjunctions, auxiliary verbs and all other function words are also placed in the high-frequency words category. Many frequent word lists in English are developed for use in pedagogic contexts, but the General Service List of English Words (GSL) of Michael West (1953) is the most commonly cited list. Although, "the GSL is in need of replacement because of its age, errors it contains, and its written focus, it is still the best available list, given a range of information it contains about the relative frequency of the meanings of the words" (Nation & Waring, 1997, p. 15). Several studies show that the coverage of the GSL is around 80% of the running words in most academic texts (Coxhead & Nation, 2001; Cobb & Horst 2004; Nation & Waring, 1997). That is, knowing the 2,000 words included in the GSL enables learners to understand 80% of the texts, or one unknown word in every five words. Learning the GSL 2,000 high frequency word families by foreign learners is an essential basis for all language use (Cobb & Horst, 2004; Coxhead & Nation, 2001; Nation & Waring, 1997).

As a criterion to select words for inclusion in the MAWLOT in the present study, we select the words outside the GSL to represent medical vocabulary.

2.5.1.2 Academic words

Academic words (vocabularies) are a group of words that is not restricted to any specific domain, and occur frequently over a wide range of academic texts across

disciplines. This means that these words are common in academic texts; however, they are not technical terms in any particular domains either. Instead, they usually occur over a range of formal papers and academic texts such as university textbooks, field specific journals, reports, manuals, etc. Some examples of academic vocabulary are *abstract, acknowledge, attitude, clarify, design, enable, implicit, job, occupy, transfer, volume,* etc. The University Word List (UWL) (Xue & Nation, 1984; Nation, 1990), and The Academic Word List (AWL) (Coxhead, 2000) are two outstanding lists of academic vocabulary.

The UWL is a compilation of four studies (Coxhead & Nation, 2001; Nation & Waring, 1997) consisting of around 800 word families not included in the first 2,000 words of the GSL. An analysis of text corpora in a variety of disciplines showed that the reliability of the GSL, together with the UWL, can increase 10% of the text coverage, i.e. from 80% to 90%. However, Coxhead and Nation (2001, p. 254) mentioned that the 570 AWL words provide better coverage of academic texts than the 800 words of the UWL. As a result, in most academic texts, knowing the GSL together with the AWL or the UWL will give coverage close to 90% of the running words. In addition, the two groups of high frequency-words and academic words "constitute a general English for Academic Purposes (EAP) vocabulary syllabus that takes a learner to the outer edge of reading in a specific domain" (Cobb & Horst, 2004, p. 319). It also was confirmed that these two groups of words "will give close to 90% coverage of the running words in most academic texts. When this is supplemented by proper nouns and technical vocabulary, learners will approach the critical 95% coverage threshold needed for reading" (Coxhead & Nation, 2001, p. 260). Therefore, the minimum lexical knowledge base for reading in any academic domain is mastery of the GSL combined with either the UWL or the AWL. This means that comprehending an academic text requires

knowing at least 3,000 word families, including 500-800 academic words in the AWL or the UWL in addition to the 2,000 high-frequency words in the GSL.

In the current study, the researcher agrees that no matter how extensively some words are used in textbooks across a range of subject areas, many of the lexical items still remain unfamiliar to students (Nation, 1990; Coxhead, 2000). In a study of typical community college textbooks, Santos (2000) found that one out of every six words, or roughly 16 percent of the words in the textbook sample, were academic words that the reader was unfamiliar with (Santos, 2000, cited in Wang, Liang, & Ge, 2008). According to Hyland and Tse (2007), vocabulary taught in schools is the vocabulary critical to understanding the concepts of the content, which they define as academic vocabulary. They also claim that in identifying academic vocabulary one should notice the degree of importance of the terms. Some terms are critically important, while some terms are useful but not critical. Some terms are interesting, but not useful. The same statements can be made for collocations and lexical phrases beyond the single word as well.

Regarding the present study, we face the lack of a medical academic word list based on medical textbooks. However, there is a medical word list, i.e. the medical academic word list (MAWL) of Wang et al. (2008), but it is entirely developed based on medical research articles, which is not enough for first-year students whose main reference materials are textbooks rather than articles. Therefore, developing a medical academic word list of textbooks (MAWLOT), as the main goal of the present study will help to fill the research gap of Wang at al.'s study in EAP, and especially in EMP.

2.5.1.3 Technical words

Words that are particularly unique and are used only in one discipline like medicine, math, or science are considered technical words. These words are discipline-specific

words (domain specific academic words or content-specific words) (Hiebert & Lubliner, 2008) that have just one meaning, the understanding of which is essential to build conceptual knowledge in the discipline. For example, the words '*tumor, anemia, hemoglobin, gonadotropin*' are restricted to medicine, whereas the words '*phoneme, morpheme, lemma, hapax legomena*' are exclusively used in applied linguistics. As can be noticed from any particular subject, technical words are probably about 1,000 words or less of the number of headwords in any specialised dictionary (Coxhead & Nation, 2001). According to Coxhead and Nation (2001) about 5% of the running words in a text typically cover the technical words. Knowledge of this 5% of words in addition to the high-frequency (the GSL) and academic words can completely enhance understanding academic texts for learners. Applying a list of technical words in a particular field besides an appropriate word list helps to increase the comprehensibility of academic texts in the field, according to Coxhead (2000). These are words that sometimes are not required to be taught by English teachers (Coxhead & Nation, 2001) and can be more naturally acquired from specialized courses.

In the current study, I have presented 53 medical technical words and examined the effect of adding them to the medical academic word list of textbooks to enhance the vocabulary knowledge of students in reading of medical texts.

2.5.1.4 Low frequency words

This group of words is so far the biggest group of words among the four groups of words and contains all the words which are not included in the other three groups.

The examples of these words, which typically occur in a very narrow range and at a low frequency, are proper nouns, words rarely found in language use, words found only once or twice in one text and seldom appearing in other texts, etc. The text coverage of this group is about 5% (Coxhead & Nation, 2001).

In a study, Carroll, Davies and Richman (1971, cited in Coxhead & Nation, 2001) showed that 40.4% of 5,000,000 running words in a corpus were 86,741 different word types occurring only once or twice. The result can certainly be used in pedagogical practice by providing the type and the number of words needed by learners to achieve their specific goals of language use. Similar to the technical words, teaching the low frequency words directly is not recommended. Teachers could prepare students to deal with these two types of words (technical words and low frequency words) by using reading strategies such as analyzing word parts, guessing word meanings from context clues, or using dictionary and word cards. Obviously, the 2,000 high frequency words of the GSL, and the UWL or the AWL should receive primary and direct attention.

As a result, it is obvious that developing a list of frequent medical academic words is necessary in EAP (EMP). Extracting a list of medical technical words may also help to enhance the comprehension of medical texts, which will be considered in the study. In addition, one criterion for word selection in the study is based on the fact that the selected words for inclusion in the medical academic word list of textbooks (MAWLOT) should be outside the GSL and the AWL.

2.6 Summary

In this chapter, English for Academic Purposes (EAP) was defined with respect to its role in university curricula all over the world, as well as considering reasons to offer an EAP course in universities. Academic vocabulary was described in terms of word classification to give a proper notion of academic word in the present study. In addition, reviewing the background studies in developing academic wordlists in EAP provides evidences of an absence of an academic word list based on textbooks in medicine. The importance of the corpus linguistics to develop specialised word lists is discussed in Chapter 3.

CHAPTER 3: CORPUS LINGUISTICS

3.1 Introduction

In this chapter, the literature related to corpus linguistics is reviewed in order to consider the strengths of the theory as a framework to develop a medical academic word list of textbooks. The review is divided into four main parts: corpus linguistics, designing a corpus, considering the stages of developing two corpora related to the study, and the notions of phraseology and collocations.

Firstly, the theory of corpus linguistics is considered to be the suitable framework for this study. Secondly, designing a corpus to develop an academic word list is the best way to derive insights for the purposes of the present study. Thirdly, the practical stages of designing the Academic Corpus of Coxhead (2000) in developing the AWL and the medical academic corpus of Wang et al. (2008) in establishing the MAWL have led me to design the Medical Academic Corpus of the present study to develop the medical academic word list based on textbooks. Fourthly, the notions of phraseology and collocations are considered to observe the words' behaviour included in the target word list (Medical Academic Word List of Textbooks) of the present study. Finally, the status of the present study in this area of research is established.

3.2 What is a corpus?

Corpus (plural: corpora) is a collection of texts that plays a role as the bodies of evidence in linguistic analysis (Thompson, 2006). According to Bennett (2010), a corpus is a "large principled collection of naturally occurring examples of language stored electronically" (p. 2). Various notions of 'corpora' have been defined by researchers. Some researchers define the corpus simply as a collection of texts (Thompson, 2006) in any format, while for other researchers the collection of texts should be only in an electronic format (Jones & Waller, 2015). In both cases a corpus is

considered a source of evidence for investigating language use in a specific context. The corpus has been also defined as "a body of written text or transcribed speech which can serve as a basis for linguistic analysis or description" (Kennedy, 2014, p. 1), or as a sample of language use (i.e. from a particular population). In addition, according to McEnery and Gabrielatos (2006) in the past forty years electronic corpora have been considered an important "resource" (p. 33) for linguists. Using electronic corpora and analyzing the data have provided us with an opportunity to access corpora information based on our research aims.

Furthermore, a corpus should act as a sample of a particular kind or a variety of discourse type, according to Thompson (2006). Francis and Kucera (1979) define a corpus as "a collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis" (p. 7). Atkins and Clear (1992) define a corpus as a body of text assembled according to explicit design criteria for a specific purpose. In EAP, four main types of EAP corpora are identified by Nesi (2104, cited in Timmis, 2015): a) corpora of expert writing, b) learner corpora, c) corpora of university student writing, and d) spoken academic corpora. Nesi (2014) states that the most common type of the EAP corpora is the corpora of expert writing, because it is available in the public domain, therefore, easy to access for the purpose of corpus compilation. The developed corpus of the present study is also the corpus of expert writing.

3.3 Types of corpora in corpus linguistics

Before compiling a corpus for the main study, we have to know which kind of corpus to develop. A corpus is often described as either a 'general' or a 'specialized' corpus (Cheng, 2011, p. 32) in corpus linguistics. General corpus usually refers to a very large corpus (often more than 10 million words) to use extensively for generalization about

language as a whole (Bennett, 2010). The most famous examples of general corpora are the Bank of English with 600 million words, Corpus of Contemporary American English (COCA) with 400 million words, and the British National Corpus (BNC) with more than 100 million words. On the other hand, specialised corpora contain texts of a certain corpus to answer specific questions, especially in an ESP setting (Bennett, 2010). This type of corpora could be large or small. However, it seems that in addition to the size of the corpus, the main factor in distinguishing these two kinds of corpora is the purpose for which they are compiled. While general corpora aim "to examine patterns of language use for a language as a whole", specialized corpora are developed "to describe language use in a specific variety, register or genre", as Cheng (2011, p. 32) mentions. Specialized corpora cover registers, genres, language forms, and language varieties. In other words, specialized corpora tend to be domain- or genre-specific (McEnery, Xiao, & Tono, 2008). For example, a specialized corpus could be in the domain of medicine (like the present study) or mathematics, and in the specific genre of textbooks (the genre of the present study), articles or newspapers. Representativeness and balance are the main factors required to select the materials for inclusion in specialized corpora.

Obviously, in the present study I have to compile a specialized corpus in the domain of medicine and in the genre of medical textbooks. In addition, before developing the medical academic corpus, I have to ensure representativeness and balance of the contents, i.e. medical textbooks for students at the general physician stage in Iran.

3.4 Defining corpus linguistics

Corpus linguistics is defined simply as compilation and analysis of corpora. However, this simple definition does not mean that everyone engaged in corpus linguistics describes it similarly. Different researchers describe corpus linguistics

differently, sometimes as a tool, a method, a methodology, and a methodological approach, or as a discipline, a theory, a theoretical approach, a paradigm (theoretical or methodological), or a combination of these.

Gries (2006) views corpus linguistics as a paradigm, but "the methodological conceptualization is favored" (cited in Taylor, 2008, p. 191). Gries mentions that "[o]ver the past few decades, corpus linguistics has become a major methodological paradigm in applied and theoretical linguistics." (p. 191). Other researchers mostly consider corpus linguistics using two main approaches. One treats corpus linguistics as a methodology to enhance research into linguistic disciplines such as lexicography, lexicology, grammar, discourse and pragmatics. The other approach, considers corpus linguistics as a discipline or theory.

Those who view corpus linguistics as a methodology use the term 'corpus-based approach'; for example, McEnry, Xiao, and Tono (2006). These researchers use corpus linguistics to test existing theories or frameworks against evidence in the corpus. McEnery, Xiao and Tono remind us that "corpus linguistics is a whole system of methods and principles of how to apply corpora in language studies and teaching/learning, it certainly has a theoretical status. Yet theoretical status is not theory in itself" (pp. 7–8). McEnery and Wilson (2001) and Meyer (2002) view corpus linguistics as a methodology as well. Bowker and Pearson (2002) also define corpus linguistics as "an approach or a methodology for studying language use" (p. 9). For these researchers corpus linguistics is an empirical method of linguistic analysis and description, using real-life examples of language data stored in corpora as the starting point (Crystal, 1992; Jackson, 2007; cited in Cheng, 2011, p. 29). They believe that "corpus linguistics is 'maturing methodologically'" (McEnry & Wilson, 2001; cited in

Cheng, 2011, p. 29) or "an approach or methodology for studying language use" (Bowker & Pearson, 2002, p. 9).

In contrast to the group who views corpus linguistics as a methodology, the other group views corpus linguistics as a discipline, like Tognini-Bonelli (2001) and Biber (2009). Tognini-Bonelli views corpus linguistics as a "pre-application methodology" that possesses "theoretical status" (p. 1). Similarly, in the viewpoint of Mahlberg (2005) corpus linguistics is "an approach to the description of English with its own theoretical framework" (p. 2). To emphasize this viewpoint Mahlberg employs the term "corpus theoretical approach" (2005, 2006). These researchers describe their approaches as 'corpus driven', and use the corpus as the starting point for developing theories about language. They believe that corpus linguistics is much more than methodology and view it as a theory (Sinclair, 1991, 1996, 2001, 2004). They also believe that corpus linguistics "re-unites data gathering and theorizing and this is leading to a qualitative change in our understanding of language" (Halliday, 1993, p. 4). They also point to the corpus linguistics as a 'bottom-up' approach that looks at "the full evidence of the corpus', analyzing the evidence with the aim of finding probabilities, trends, patterns, co-occurrences of elements, features or grouping of features" (Teubert & Krishnamurthy, 2007, p. 6), which would also explain corpus linguistics as 'a new philosophical approach to linguistic enquiry' (p. 2).

Similarly, Stubbs (2001) views corpus linguistics as a theory, and writes in a note to Sinclair (1991) that "[i]n this vision of the subject, a corpus is not merely a tool of linguistic analysis but an important concept in linguistic theory" (pp. 23 – 24), and as a discipline in its own right (Cheng, 2011). He also mentions that: "Corpus linguistics is not concerned with what happens to occur (at least once). It is concerned with a much deeper notion: **what frequently and typically occurs**" (p. 151, emphasis mine).

Answering these questions mostly leads researchers to develop word lists based on corpora, as in the present study.

Teubert (2005) describes corpus linguistics as "a theoretical approach to the study of language" (p. 2), and emphasizes the theoretical conceptualization. Leech (1992) also argues that "computer corpus linguistics defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject" (p. 106).

On the other hand, according to Taylor (2008), in several articles researchers have used corpus linguistics both as a methodology and field, which may be explained by the transversal nature of corpus linguistics, or as Partington mentions, "by the way in which the discipline is driven by the technology" (cited in Taylor, 2008, p. 196). Partington states that:

"To make such clear distinctions between instruments and enterprise is anachronistic. Firstly, because observation informs theory just as theory informs observation […]. Secondly, in most modern physical science, the object of observation is only tangible, in a sense only exists for an observer (outside a mathematical formula), through the instruments of study, which thus constrain not only what can be perceived but even the very questions that can usefully be put of the physical world" (Partington, cited in Taylor, p. 196).

In an interesting part of her study, Taylor also considers the concordance lines of the web of the sequence *corpus + linguistics + is* to collect examples by using WebCorp10. She reports that "While corpus linguistics is often discussed in terms of method, methodology, and also technology and tool, whether it is viewed as 'possessing' these functions or "being" one of them is often unclear. Also, corpus linguistics has never

been described as "being" a theory in itself, maybe because it has developed its own theories of language study" (p. 195).

Regarding the present study, however, I agree that "the debate is likely to continue, and may never be fully resolved" (Cheng 2011, p. 30), but I view corpus linguistics as "an approach to description of English with its own theoretical framework" (Mahlberg, 2005, p. 2) to develop and describe the medical academic word list of textbooks (MAWLOT). I also analyse the obtained evidence of the Medical Academic Corpus of the study to find probabilities and co-occurrences of elements of the corpus, as Halliday (1993) and Teubert and Krishnamurthy (2007) mention.

In addition, similar to Nation (2001) and Cobb and Horst (2004), who are interested in learning words in instructional contexts and thus concentrate more on word frequency and their corpora collocations while showing less interest in the demonstration and processing of richer semantic representations, my focus in this study is also on describing the frequency of words and their corpora collocations, not to process the semantic representations of the words selected for inclusion in the medical academic word list of textbooks.

3.5 Corpus design

A variety of corpora based on different disciplines were established and analyzed in order to develop different kinds of word lists to improve the learning efficiency of learners of academic vocabulary in academic domains (Hajiyeva, 2015). Designing a corpus to get information related to the study's aims needs to meet some criteria. Generally, a corpus usually refers to a collection of texts and is defined as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible" (Sinclair, 2004, p. 16). A corpus is also "a collection of texts assumed to be representative of a given language put together so that it can be used

for linguistics analysis" (Tognini-Bonelli, 2001, p. 2), or "collection of texts (or parts of text) that are stored and accessed electronically" (Hunston, 2002, p. 2) for linguistics study. It also should be reminded that a corpus does not refer to any collection of texts; it is "a collection of naturally occurring language texts, chosen to characterize a state or variety of language" (Sinclair, 1991, p. 171).

The above sentences reveal that a corpus is designed and compiled based on corpus design principles. One of the corpus design principles is 'orientation', as Sinclair (2004) mentions. That is, "only those components in the corpus that are designed to be independently contrasted are contrasted" (p. 3). In addition, designing a corpus needs to consider the representativeness and balance of the corpus to maintain adequate coverage. In addition, "External, rather than internal, criteria need to be used when selecting subject matter for the corpus" (p. 11). External criteria are defined situationally (e.g. genres, registers) and refers to: the mode of the text (e.g. speech, writing and electronic mode); the type of text (e.g. in designing a written corpus, whether books, newspapers, journals, or letters should be selected); the location of the texts (e.g. the English of the UK or USA); the domain of the text (e.g. popular or academic), etc. Internal criteria are defined linguistically (text types) and are used to define representativeness, e.g. distributions of words or grammatical constructions. This distributional argument is problematic for some researchers, since it pre-defines word distribution while it is what the corpus should tell us. For instance, Berber-Sardinha (1993, cited in McEnery, Xiao, & Tono, 2008) mentions that a representative corpus should include the majority of the word types in the language as recorded in a comprehensive dictionary. This definition shows that researchers employing internal criteria expect that the corpus will present particular words sufficiently, in spite of the fact that this is what the corpus itself should tell us.

In the viewpoint of the researcher of the present study, external criteria seems more justified, since linguistic characteristics are independent of the selection process, while internal analysis can just help to define what texts should next be added.

3.5.1 Representativeness in corpus linguistics

Representativeness refers to the extent to which a sample includes the full range of variability in a population (Leech, 1992). However, according to Chomsky (1975) there is no such thing as a representative corpus, it has not been very well defined (Kilgarriff & Grefenstette, 2003, p. 343). Generally, there is no rule for developing a corpus (Coxhead, 2000), but an issue like representativeness should be considered in developing any corpus (Biber, 1993) to allow findings to be generalised to a particular variety of language.

Most researchers focus on the sample size as the most important factor in achieving representativeness by considering how many texts must be included in the corpus, as well as how many words per text sample, but, a great deal of research on sampling theory confirms that the most important consideration in achieving representativeness sample is not the size of the sample; rather, it is the target population which enables researchers to achieve a representative sample (Biber, 1993). According to Biber (1993), representativeness is determined by balance (i.e. the range of genres in the corpus) and sampling (i.e. how text chunks for each genre are selected).

3.5.2 Balance

Representativeness depends upon how balanced the corpus is, but some researchers believe that the notion of balance is even more ambiguous than representativeness. A balanced corpus is defined as a corpus (general or specialized) which covers a wide range of text categories, and designing it needs to include a great variety of language samples. There is no instrument to measure the balance of a corpus, and it is usually

measured based on proportional sampling. The important different axes in designing a balanced corpus are: written and spoken language, genres, production variables (gender, age, social class, etc.), reception variables (large audience, small audience, level of formality, etc.), and temporal slices. A specialised corpus such as that of the present study, which tends to be domain- or genre-specific, should be balanced by including a wide range of types to claim the representativeness of a particular variety of language. Achieving representativeness and balance in designing a corpus needs to sample language that is representative of the larger population.

3.5.3 Population

Definition of population is the first concern in designing any corpus, which covers at least two aspects: (1) the boundaries of the population (i.e. considering what texts should include or excluded from the population); and (2) hierarchical organisation within the population (i.e. considering categories of the texts included in the population and their definitions) (Biber, 1993). Before collecting the samples, there should be enough attention devoted to these concerns in designing text corpora. Two good examples of well-defined corpus populations are the Brown corpus (Francis & Kucera, 1979) and the LOB corpus (the Lancaster- Oslo/Bergen) (Johansson, 1980). Both target populations in these corpora have been defined based on their boundaries, i.e. they include all published English texts printed in 1961 in the United States and United Kingdom, respectively. In addition, they have been organised hierarchically, i.e. they include fifteen major text categories and numerous subgenre distinctions within these categories. The other issue to consider in designing a representativeness corpus is having a good sampling frame, which can be also seen in these two corpora.

In the present study, the target population is the published English medical textbooks printed as the last references in 2011 (the time of starting the study) in subdisciplines of medicine for general physicians (GPs) in Iran.

3.5.4 Sampling frame

Having an adequate sampling frame provides the possibility of selecting a probabilistic sample. “A sampling frame is an operational definition of the population, an itemized listing of population members from which a representative sample can be chosen” (Biber, 1993, p. 244). Considering the LOB corpus manual (Johansson, 1980), it appears that the sampling frame for books is all 1961 publications listed in The British National Bibliography Cumulated Subject Index, 1960-1964, based on the subject divisions of the Dewey decimal classification system. The target population for periodicals and newspapers is all 1961 publications listed in Willing's Press Guide (1961). The sampling frame in the case of the Brown corpus is the collection of books and periodicals in the Brown University Library and the Providence Athenaeum - which is however, less representative of the total texts in print in 1961 than the Lancaster-Oslo/Bergen (LOB) corpus, but it does provide well-defined boundaries and an itemized listing of members.

Selecting a probabilistic sample is also important to yield an adequate sampling frame. There are a lot of probabilistic samples constructed based on random selection and an equal chance of being selected in the population, but the method employed in constructing the Brown and LOB corpora is stratified sampling (Biber, 1993). In this method, after identifying subgroups within the target population (in the case of the Brown and LOB corpora, the genres), each of those “strata” (p. 244) are sampled with random techniques. Biber believes that employing this approach guarantees adequate representations of all strata, as well as selecting a non-biased sample within each

stratum at the same time. These are the reasons for researchers to consider the Brown and LOB corpora as the 100% representative (in genre categories) and unbiased corpora for selecting texts within each genre.

In addition to the above criteria in designing a corpus, three main criteria, i.e. representation, organization, and size should be espeifically considered to design a corpus in developing a word list. The way that words are defined for inclusion in the word lists is also an important criterion to consider in developing any word list, as well as the medical academic word list of textbooks (MAWLOT).

3.6 Word selection

Having a proper definition of word is an important issue in developing a word list (Coxhead, 2000). Researchers believe that the importance of word selection's criterion origins in the fact that different criteria lead to different results. To develop the AWL, Coxhead preferred to define the word as the unit of a word family based on the definition of Bauer and Nation (1993). According to Bauer and Nation, a word family is a stem and its inflected forms and derivations. That is, a word family includes all closely related affixed forms as well as the most frequent stem with its regular and productive prefixes and suffixes, in addition to the perceived transparency. Based on this definition, the affixes are the only ones that "can be added to stem and can stand as free forms" (Coxhead, 2000, p. 218). Therefore, for example, the words *specify* and *special* would not be counted in the same word family because of *spec*, which is not a free form. But, the words *acid*, *acids*, *acidly*, *acidic*, and *acidity* are considered to be parts of the same word family. Furthermore, researchers believe that "in the mental lexicon word families are an important unit" (Nagy et al., 1989, p. 262). Coxhead (2000) argues that lexical items that may seem morphologically distinct from one another are sufficiently related to count as a single lexical item. She reminds us that

English learners prefer word lists which usually group words into families because "comprehending regularly inflected or derived members of a family does not require much more effort by learners if they know the base word and if they have control of basic word-building processes" (p. 218). In addition, according to Nation and Webb (2011) the most sensible unit when counting for receptive knowledge, i.e. for reading and listening, is the word family. Hajiyeva (2015) states that the reason to use the word family as the unit of counting is that "if one or two members of the family are known, then little learning is required for receptive use (comprehension) of their family members" (p. 141).

Furthermore, according to Coxhead two methods have been employed by researchers to select words for inclusion in academic word lists. The first method is the method of Lynn (1973) and Ghadessy (1979). Lynn and Ghadessy noted annotations as a signal of student difficulties in comprehending texts, and selected the words based on the students' annotations in textbooks. The second method was employed by Campion and Elley (1971) based on the occurrence of words in 3 or more subject areas out of 19. Campion and Elley also included the degree of native speakers' familiarity to select the words in the word list. However, the number of running words in their corpus was too small (301,800 running words) to meet the initial criteria (Coxhead, 2000). Praninskas (1972) included the criterion of range to develop the word list. Similar to the study of Campion and Elley, the corpus of Praninskas was small in the number of running words (272,000 running words), and the resulting list was also small without much variety in the words.

In the present study, I use the definition of Bauer and Nation (1993); i.e. the word for inclusion in the Medical Academic Word List of Textbooks (MAWLOT) is selected as a unit of a word family (a stem and its frequent inflected forms and derivations). In

addition, the process of creating other corpora to develop academic word lists guides me to create the medical academic corpus of the present study. Therefore, two famous corpora related to the present study — i.e. the Academic Corpus of Coxhead (2000) and the medical academic corpus of Wang et al. (2008) — are considered as samples of corpora to develop the Medical Academic Corpus of the present study.

3.7 Developing the corpus of the AWL (Coxhead, 2000)

The corpus created by Coxhead (2000) contained 28 subject areas, organized into 7 general areas and within four disciplines of arts, commerce, science and law. Developing the corpus involved collecting each text in electronic form, eliminating its bibliography, and counting its words. After balancing the number of short, medium-length, and long texts, Coxhead inserted each text into its subject area. According to Coxhead, balancing the subcorpora is essential to organizing the corpus. Therefore, four main subcorpora of her study (arts, law, commerce and science) with approximately equal numbers of running words were included in the corpus. That is, each subcorpus contained approximately 875, 000 running words or 3.5 million running words.

The Academic Corpus was organised to permit the range of occurrence of particular words to be examined, as Coxhead mentions. Besides, the majority of the texts had been written for international audiences. Sixty-four percent was sourced in New Zealand, 20% in Britain, 13% in the United States, 2% in Canada, and 1% in Australia. The Academic Corpus was analysed by the corpus analysis program, i.e. *Range* (Heatley & Nation, 1996) to count and sort the words in the Academic Corpus. The software of Range was the basic program to design the software program of Complete Lexical Tutor (CLT) by Cobb et al. (2005) for corpus-based studies.

3.8 Developing the Academic Word List (AWL)

According to Coxhead (2000), three key principles guided her to select words for inclusion in the Academic Word List. These three principals were: specialised occurrence, range, and frequency.

1. Specialised occurrence: The words included in the AWL had to be drawn from the 2,000 most frequent words in West's (1953) General Service List of English Words (GSL). Coxhead employed the principle of specialised occurrence to select only academic words which were not among the frequent general English words of the GSL provided for non-native English speaking students.
2. Range: A member of a word family included in the AWL had to occur at least in 15 or more out of the 28 subject areas (more than half of the subjects).
3. Frequency: The words included in the AWL had to occur at least 100 times or more in the Academic Corpus.

Frequency was considered secondary to range in developing the AWL to avoid the bias of including large texts in the corpus. According to Coxhead (2000), the *Collins COBUILD Dictionary* (1995), which had been collected based on the corpus of newspapers from the early 1990s, had highlighted *Yemeni* and *Lithuanian* as the high-frequency words. These words are not among the most frequent words in the English language in the GSL, but they were extensively used by newspapers at that time. Coxhead strictly employed a frequency of 100 for multiple-member word families in the AWL, but not for single-member families. According to Coxhead, studying the Brown Corpus (Francis & Kucera, 1979) shows that a corpus of around 3. 5 million running words would identify 100 occurrences of a word family. The words like *analyse* and *concept* are among the most frequent word families in the AWL. Some of the least frequent words of the AWL are *convince* and p*ersist*.

Finally, Coxhead included 570 word families in the AWL. According to Coxhead, the AWL's word families occurred in a wide range of the subject areas in the Academic Corpus. The coverage of the AWL was 10% of the vocabulary in the Academic Corpus. For disciplines of arts and law the coverage was 9.3% and 9.4%, and for science 9.1%. The highest coverage belonged to commerce with 12.0%. That is, the list could therefore be more advantageous for commerce students. The Academic Word List (AWL) has been widely tested and used by scholars, teachers, and material developers for more than 10 years in English for Academic Purposes (EAP) (Coxhead, 2010).

3.9 Developing the corpus of the MAWL (Wang et al., 2008)

To establish the corpus of medical academic word list based on medical research articles (RAs), Wang et al. (2008) selected a written specialised corpus containing 1,093,011 running words from the written texts of 288 medical research articles. The data to include in the corpus of written medical RAs was retrieved from the database Science Direct Online, the world's largest electronic collection of science, technology and medicine articles (Wang et al., 2008). Wang et al. included the articles with their original length, i.e. the full text as well as the bibliographic information. The texts were selected from 1800 journals in 24 disciplines from natural science to social science, medicine and dentistry in 32 subject areas at their time of the research. The articles had to be written in the internationally conventionalized IMRD (Introduction-Method-Result-Discussion) structure. The articles had also been published in the years of 2000-2006, written only by native English speakers, which affiliated with an institute in the countries for which English was the first language.

According to Wang et al. (2008), the articles were selected for inclusion in the corpus in three rounds. First, the researchers considered each of the 32 subject areas as a layer, and then selected 3 journals from each layer, totaling 96 journals. Second, from

96 journals one issue was randomly selected. The articles that were not written in the IMRD format, were shorter than 2000 running words, were longer than 12,000 running words, or had not been written by native speakers were excluded. Third, three criteria-fulfiling articles were selected by simple random selecting from each of the 96 issues. Finally, Wang et al. selected 288 texts of medical RAs to include in the corpus. The shortest text contained 2923 running words, and the largest one contained 10,901 running words. On average, each text contained 4939 tokens to include in the corpus of Wang's et al. MAWL.

3.9.1 Data processing

Processing the data to establish the medical academic word list based on the research articles was done in two stages: the standardisation of medical RAs for storing in the corpus, and the normalization of words in the RAs stored in a corpus (Wang et al., 2008). To standardise the medical RAs included in the corpus, all parts that were not readable by computer analysing programs to process like charts, diagrams, and bibliographies were removed. The second stage, i.e. the normalization of words, was done by the computer software which was able to read all inflections and derivations of a word as well as its basic form.

Similar to Coxhead (2000), the definition of word family to establish the Wang's et al. MAWL was also the definition of Bauer and Nation (1993), i.e. stems plus all closely related affixed forms.

3.9.2 Word selection

To establish the MAWL, Wang et al. (2008) followed the method of Coxhead (2000) in developing the AWL. That is, for the criterion of range Coxhead selected the word families from those for which their members occurred in more than a half of the 28 subject areas in the corpus. Wang et al. also set 50% as the criterion of range to include

the word families; i.e. the members of word families had to occur at least in 16 out of 32 subject areas of the corpus. The number of running words in Coxhead's Academic Corpus was approximately 3,500,000, and the number of running words in the corpus of Wang et al. was 1,093,011 (a third of Coxhead's running words in the Academic Corpus).

The lowest frequency of word family members to be counted in the MAWL was 30 times, i.e. a third of 100 times, which Coxhead (2000) employed to develop the AWL as the criterion of frequency. Similar to Coxhead's method of word selection, Wang et al. applied the criterion of frequency as secondary to range. According to Coxhead (2000), counting a word in a word list based only on frequency will cause a bias toward longer texts and topic-related words. In the Wang's et al. MAWL, just the word families which covered 16 or more subject areas out of 32 were included in the MAWL. The words that occurred with high frequency but covered fewer than 16 subject areas were excluded from the MAWL.

In sum, the criteria of word selection for developing Wang's et al. (2008) MAWL were three principles:

1. Specialised occurrence: The included word families had to be outside the West's (1953) *General Service List of English Words* (GSL), i.e. the first 2000 most frequent word families of English.
2. Range: The word families had to occur in 16 or more subject areas out of 32.
3. Frequency: The word families had to occur at least 30 times in the corpus of medical research articles.

3.9.3 Developing the Medical Academic Word List (MAWL)

In developing the MAWL based on medical RAs after standardization of medical RAs and normalization of words to include in the corpus, the frequency and range of

word families were counted by the computer software (Wang et al., 2008). The word families which had occurred in at least 16 or more subject areas out of 32 in the corpus were selected. These word families also had to occur at least 30 times in the corpus of medical RAs (a third of Coxhead's 100 times), because the number of the running words in their corpus (1,000,000) was about one-third of Coxhead's corpus (3,500,000), as Wang et al. mentions.

The MAWL contained 623 word families, which accounted for 12.24% of tokens (running words) in the corpus of medical RAs. In the final list, *cell* was the most frequent word in all 32 subject areas of the corpus with 4421 occurrences, and the word *static*, occurring only 30 times in 20 subject areas of the corpus, was the least frequent word of the MAWL.

In the present study, with focus on developing a medical academic word list of textbooks, we have to design a balanced corpus with at least 3.5 million running words consisting of texts of major fields in medicine for students at the general physician stage, to represent medical academic vocabulary. We also select the *words* through the unit of word family based on the definition of Bauer and Nation (1993), i.e. a stem as well as its inflected forms and derivations.

In addition to developing the medical academic word list of textbooks, the corpus information helps us to retrieve the collocations of words. To retrieve the collocations of the most frequent words included in the MAWLOT, we have to consider some notions related to collocations such as collocation, concordance and concordancer, as well as the basic concepts of phraseology.

3.10 Phraseology in EAP

Studying research in EAP appears that the current trends in EAP research are very much like those observed in other areas in terms of both the analysis of discourse for specific purposes (e.g. professional language-in-use) and the teaching of general English. To start with the latter, Coxhead (2010) has written a whole chapter on the topic of phraseology in EAP, outlining both challenges and opportunities. The change from single words to phrases or prefabricated formulaic multi-word expressions (Sinclair, 1991) has been ongoing since the 1970s, but has come into its own only much more recently, thanks largely to the corpus revolution.

Phraseology, as the central element of corpus linguistics, has been defined simply as the study of phrases (Bennett, 2010). The definition of phraseology has an important place in linguistics, because the meaning of a word is understood through phrases and several words in a sequence (Sinclair, 1991). Gläser (1988) defines phraseology, as "the linguistic description of set expressions whose meaning cannot be derived from the meaning of their parts" (p. 265). However, in the viewpoint of McKenny (2006), this definition presents phraseology just as the study of idioms and restricted collocations.

To quote Coxhead, "No word is an island" (2010, p. 149). It is important to know about words or keywords in context, and to use that knowledge to meet the language needs of students the world over. However, it is one thing to rely on corpus research to identify what collocations to teach and what pedagogical approach to adopt; it is quite another to improve students' collocational fluency in both writing and speaking. Not surprisingly, given its connection to phraseology, a second major strand in EAP is corpus analysis. This topic has also been discussed by Coxhead in the same year; namely, the usefulness of corpus analysis for EAP (Coxhead, 2010) and earlier by

Thompson (2006), whose study concludes that the integration of corpus remains as yet very restricted in the EAP classroom.

Several studies have mentioned the existence of an EAP-specific phraseology characterized by word combinations, fundamentally semantically and syntactically compositional, such as: *the extent to which*, *it is likely that*, *in the presence of*, *the aim of the study*, etc. (Biber, 1993). These studies have also indicated that beginners and professional writers differ in their use of EAP-specific lexical packages (Cortés, 2002).

Finally, in spite of different definitions and viewpoints in defining phraseology, Bennett (2010) believes that understanding the depth concept of phraseology requires an understanding of the concepts of collocations, which is discussed in the next section.

3.11 Collocation

According to Wu et al. (2016), collocations are one of the most challenging aspects of language learning, which their importance for success in language learning are widely recognised. Firth (1957), who is known as the father of collocation, quoted "You shall know a word by the company it keeps!" (p. 11). Firth invented the term collocations for usual word combinations such as" *strong accent"* but not *"powerful accent*". Or, one can say "*heavy rain"* but not "*strong rain*" (Firth, 1957, cited in Reza & Ashouri, 2016). The combination of word should be also dependent on a "particular word choice", with a "clear meaning associated with it" (Hunston & Francis, 2000, p. 37).

It is also mostly approved that collocations play a crucial role in second language learning (Lewis, 2000). According to Ackerman and Chen (2013), " [lexical collocations] contain certain variability and are thus more dynamic, while grammatical collocations or idioms consist of comparatively fixed patterns and are consequently more predictable. The former is more challenging for learners to master whereas the

latter can generally be treated as holistic units and learners can more easily internalize the usage into their lexicon" (p. 246).

In addition, there is a reasonable degree of agreement among researchers about the definition of collocation in describing it as a combination of words, which presents the evidence of specialisation of meaning in the researchers' viewpoints (McKenny, 2006). Some researchers have argued that different types of text reflect different collocation patterns. For example, Partington (1998) argues that, "Collocational normality is dependent on genre, register and style" (p. 17). Similarly, Lewis (2000) believes that "different kinds of text have radically different collocational profiles" (p. 186). Murison-Bowie (1996, p. 194) also states that the different meanings of *heart* in romantic and medical texts are indicated clearly by the different collocations (*Heartthrob*- romantic novel; *heart failure* – medical text). Gavioli (1997, pp. 87-88) also notes the differences for the most frequent collocates of *criminal* in newspapers and academic texts of social science. Her analysis of an English newspaper corpus reveals the most frequent collocates of the word *criminal* are *war/act/law*, while analyzing the word in a corpus of academic texts appears *law/liability/English* as the most frequent collocates of *criminal*.

Furthermore, the term 'collocation' has been earlier used by Palmer (1933) in Teaching English as a Foreign Language (TEFL). According to Palmer: "A collocation is a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts" (p. 1). Cheng (2011) also mentions that "the collocates of a word are words which are most frequently co-selected in the vicinity of that word" (p. 77). He goes on to state that "If two words collocate significantly, their presence in the text is the result of a single choice, not two" (p. 77). In the viewpoint of Bennett (2000), collocation is a prominent way in studying phraseology to find the

statistical tendency of words to co-occur. For example, the word *deal* is used, mostly, with *great* or *big* (*great deal* or *big deal*). Applying each word with *deal* has its own significant meaning, such as '*a great deal'* refers to a quantity, but '*a big deal*' is used to refer to an event or situation. The redundancy created by occurrence of words in texts allows readers to concentrate on the writing content rather than writing manner, because most of the time people just want to get their message across by being conventional, predictable, and unremarkable in their language, rather than impressing the audience by their use of language (Fox, 1998).

Sinclair (1991) believes that collocation is "the occurrence of two or more words within a short space of each other in a text" (p. 1). For Kjellmer (1984) collocation is also a combination of words that occurs more often in the corpus of the constituents. Kjellmer believes that these combinations must be grammatical in order to be counted as a collocation. He also suggested that in a corpus (like the Brown corpus, with one million words) a string of words must happen more than once. The importance of using collocations in language teaching has been considered by Lewis in 2000. Lewis emphasized on the role of word combination patterns in language and their use in teaching and learning a language. He also believed that collocations provide a more practical approach to language teaching syllabus design. In addition, Lewis (2000) counted collocation as the quickest path to acquiring the elements of learning a word, such as syntactical, phonological, and form information. Hsu & Hsu (2007) have examined whether the direct teaching of lexical collocations boosts general fluency of learners. The results of their study illustrated the positive impact of collocation on the learners' language skills. McCarthy (1990) stated that there is a great significance of collocations in vocabulary acquisition. He also mentioned that the association of collocation is essential in the learning vocabulary, and collocation is a central organizing code in the vocabulary of any language.

In addition, according to Howarth (1998) empirical studies on the understanding of collocations reveal that different groups of ESL or EFL learners meet special problems in producing appropriate word combinations because of lack of collocation knowledge. Lehecka (2015) counts collocation as a fundamental concept in usage-based studies in many linguistic fields, most notably lexical syntax and semantics. Lehecka also states that studying collocations in large electronic corpora allows for statistical analyses of the co-occurrence patterns of linguistic items.

In the current study, we also consider the collocation/phraseological patterns of the most frequent words included in the medical academic word list of textbooks to observe the words' behaviour in their context.

3.11.1 A concordance and concordancer

A *concordance* has been defined as "an alphabetical listing of words in a text or collection of texts, together with the contexts in which they appear" (Godwin-Jones, 2001, p. 8). In other words, a *concordance* is a list of occurrences of either a particular word, a part of a word, or a combination of words in context. An occurrence of a particular word is usually called a *keyword*, which is utilised by a *concordancer*.

A *'concordancer'* is a computer program for working with language corpora to search, access and analyse corpus information, and then to display the output in concordance lines. A concordancing software program is able to make a concordance list to show the contexts of every occurrence of a selected word or phrase in a text corpus. This means that using a typical concordancer allows us to enter a word or phrase and search for multiple examples of how that word is used in the corpus. Obviously, it is particularly useful to exploring the relationships between words by providing accurate information about the way language is authentically used. Collocation patterning and revealing different senses of a word type are the useful information provided by

concordancer software. The relative frequency of different uses of a word type is the other information that can be calculated by the software. To access corpus information by using a concordancer, concordances can be produced in a number of formats, most usefully the *Keyword in Context (KWIC)* format. The *KWIC format* displays the keyword in the center of the line with more contexts on each side of the keyword, and each occurrence of the word is listed on a separate line. The sequences of contexts can be displayed on the left or right of the keywords to give a picture of the environments where a keyword occurs in a corpus.

Regarding the present study, we face an absence of a concordance list in medicine. Having a concordance list of the words included in the word list will help us to study the words in their contexts. It also reveals the word sequences (prefabs) of the most frequent words included in the medical academic word list of textbooks (MAWLOT), which cannot be expressed by a sole word.

3.12 The present study

Based on the literature review in Chapter 2 and Chapter 3, there is no empirical evidence concerning a medical academic word list based on just textbooks. Therefore, there is a research gap in this area. The present study is distinctive from other previous studies in at least four respects.

Firstly, it has a different aim in developing a word list from textbooks for medicine only. In previous works, researchers attempted to develop word lists in other academic areas, such as general English, applied linguistics, engineering, mathematics, etc. In medicine, however, very little work has been done in developing medical word lists. For example, Chen and Ge (2007) conducted a lexical study on word frequencies and text coverage (range) of the 570 word families of the AWL in a corpus of medical research articles. But, only Wang et al.'s (2008) study was implemented to establish a medical

academic word list (MAWL) with focus on online research articles. The corpus used in the study of Lei and Liu (2016) is also a combination of research articles and textbooks, in contrast with the present study which attempts to develop a medical academic word list just based on textbooks.

Secondly, previous studies in developing a word list in medicine have been based on a corpus with a small size. For example, the corpus of Chen and Ge (2007) was compiled from 190,425 running words taken from 50 medical research articles written in English. The corpus of Wang et al. (2008) in developing a medical academic word list based on the research articles (MAWL) was also compiled from 1,093,011 running words. In contrast, the corpus of the present study contained around 3. 5 million running words from medical reference textbooks to develop a medical academic word list based on those textbooks.

Thirdly, previous studies in developing medical word lists have presented only a list of frequent medical words (based on research articles) without considering collocations and prefabs of the words. In contrast, the present study presents collocations of the most frequent words included in the medical academic word list of textbooks (MAWLOT) to facilitate the process of learners' L2 fluency by determining which words form high-frequency, multi-word phrases (Schmitt, 2004).

The fourth respect is that the present study offers a list of medical technical words in addition to their collocations to increase the students' vocabulary knowledge through reading medical texts. There has been no attempt to formulate a medical technical word list in previous studies, even in the study of Wang et al. (2008), which was conducted to establish a medical academic word list.

To summarise, the present study attempts to bridge the gaps in this area of research. Developing a medical academic word list based on textbooks can provide EAP and EMP teachers and medical students with a reference list of words to study which have occurred frequently in a wide range of medical textbooks. These words, as well as the list of medical technical words and their collocations, make up the findings of the study and should contribute to the area of English for Academic Purposes (EAP) in providing useful pedagogical implications as well as empirical evidence in the areas where research is lacking.

3.13 Summary

In this chapter, corpus linguistics was utilised as the main theoretical framework of the study in designing a proper corpus to develop the medical academic word list based on textbooks. Considering the general factors in designing a corpus beyond the reviewing stages in the case of two corpora, i.e. the Academic Corpus of Coxhead (2000) in developing the AWL and the medical academic corpus of Wang et al. (2008) in establishing the MAWL, has given me enough examples to design the corpus of the study. Reviewing the notions of phraseology and collocations have also shown me how I can make a better contribution in EAP by presenting a concordance of the most frequent words included in the target word list of the study.

Therefore, the present study is an attempt to develop a medical academic word list based on textbooks through a corpus-based lexical study. It also aims to produce the collocations of the words because learner's L2 fluency depends to a large extent on his or her understanding and usage of collocations, i.e. multi-word expressions, as Schmitt (2004) mentions.

CHAPTER 4: METHODOLOGY

4.1 Introduction

This chapter describes the research methodology of the study. The description begins with recognizing the type of corpora needed to compile the main corpus of the study, population and sample, and stages of research. Next, the pilot study is briefly described with results and suggestions for improving the main study. The description of materials and instruments is divided into three main topics: the method of corpus compilation, word selection, and research instruments. Regarding the corpus compilation, the chapter describes the method to develop the Medical Academic Corpus. The chapter also explains the method of word selection so as to show how the target word list of the study is constructed.

4.2 Population and samples

4.2.1 Population

As stated in the previous chapter, definition of population is the first concern in designing any corpus. The boundaries of the population (i.e. what texts should be included or excluded from the population), and the hierarchical organization within the population (i.e. considering categories of the texts included in the population and their definitions) are two important aspects that should be considered in defining population when designing a corpus (Biber, 1993).

Therefore, the target population in the present study is the published English medical textbooks printed as the last references for students in 2011. This does not mean that all selected textbooks were published in 2011, but they might have been published before 2011 and were used as the last references for students in 2011. For example, the *Cecile Essentials of Medicine* is the reference textbook of internal medicine that is published every 5 years. At the time of this study started, the last edition of the textbook was the

8th edition, 2010. The selected textbooks belong to major disciplines of medicine used by English-speaking medical universities considered in the study, which are considered as the reference textbooks in Iran for medical students at the stage of general physicians (GPs).

4.2.2 Samples

To develop the corpus, I first retrieved subject areas from all disciplines of medicine. The system of medical training in Iran is close to the American system; therefore, we retrieved the subject areas from universities considered in Iran to be English-speaking medical universities, such as Harvard, UCLA, University of Washington, Dalhousie University, University of Kansas, University of Dundee (UK), and University of Toronto. At the well-known medical universities in Iran such as the University of Tehran, the University of Shiraz and the University of Mashhad, these textbooks are being used as the reference texts for general physicians as well.

Examination of the medical universities' curricula reveals that some subject areas are counted as the major subjects in medicine and divided into smaller subjects (see Table 4.1). These major subjects (fields) are internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery. For example, internal medicine as a major subject covers neurology, cardiology, gastroenterology, or, obstetrics and gynecology which covers obstetrics, gynecology and women's health.

Table 4.1: Subject areas of medicine according to the Science Direct (cited in Wang et al., 2008)

Subdisciplines	Covered subjects
Internal medicine	Anesthesiology and pain medicine Cardiology and Cardiovascular Medicine Clinical Neurology Nephrology Critical Care and Intensive Care Dermatology Emergency Medicine Endocrinology, Diabetes and Metabolism Gastroenterology Hematology Hematology Immunology, Allergology and Rheumatology Infectious diseases Oncology Ophthalmology Pulmonary and Respiratory Medicine Urology
Obstetrics & Gynecology	Obstetrics Gynecology Women's health
Pediatrics	Perinatology Pediatrics Child Health
Psychiatry	Psychiatry Mental Health
Surgery	Orthopedics Sport Medicine Rehabilitation Otorhinolaryngology Facial Plastic Surgery Surgery

In the current study, the main goal of selecting the medical textbooks is based on the fact that the subjects should represent (as much as possible) the reading of medical students at the general physician stage. Therefore, further analysis of the programs also revealed the latest reference (s) for the various major fields. For example, *Cecile Essentials of Medicine* (2010) was selected as the reference for internal medicine in our study. This textbook was selected since it covers all subjects within the major discipline of internal medicine. Also, it has enough words to consider the data of the internal medicine subject of the study. The textbook selected for psychiatry, i.e. *Kaplan &*

Sadock's Synopsis of Psychiatry: *Behavioural Sciences/Clinical Psychiatry* (10th Ed., 2008) also contained more than sufficient needed words. However, for other major fields, i.e. obstetrics and gynecology, pediatrics, and surgery I had to consider more than one reference textbook to collect the needed data (see Appendix A and Table 4.4 for the selected reference textbooks to include in the Medical Academic Corpus). All the selected textbooks are known as the main reference for their respective fields in medical universities in Iran and many other countries. Furthermore, the portion of each subdiscipline, as mentioned in the literature review, should be balanced and coherent. This means that in the Medical Academic Corpus of the current study with 3. 5 million running words, each subcorpus has to contain around 700,000 running words.

4.3 Stages of research

A pilot study was implemented to test the samples and instruments. Next, the main study was conducted and the data was analysed and interpreted and the findings were reported.

4.3.1 Pilot study

Before the main study was conducted, a pilot study was implemented to test the samples and instruments as well as to assess problems or difficulties which might arise during the study. The implementation of the pilot study was expected to obtain insights and guidelines of the selected samples and instruments for the main study, because developing a medical academic word list based on textbooks was new to vocabulary word list research. In addition, problems concerning the use of the samples and instruments could be identified and tackled before the implementation of the main study.

Pilot study is defined as a feasibility study intended to guide the planning of a large-scale investigation (Thabane et al., 2010). That is, pilot studies are intended to provide

useful information about the feasibility of the main study (Thabane et al., 2016). According to Eldridge et al. (2016), there are a number of definitions of pilot studies presented by researchers (e.g. Thabane et al., 2010), which most have in common the idea of conducting a study in advance of a larger, more comprehensive, investigation. Eldridge et al. (2016) also mention that the feasibility should be the main emphasis of a pilot study and agree with the opinion that a pilot study is synonymous with a feasibility study intended to guide the planning of a large scale investigation. Therefore, this section describes the process of conducting the pilot study to identify the adequate corpus size, besides to examine the instruments of the study to conduct the main study.

To compile a corpus for the pilot study, a small written specialized corpus of around 400,000 running words (tokens) was developed from the identified sub-disciplines, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery. The samples also consisted of textbooks similar to those in the main study. For example, the selected texts for the subcorpus of internal medicine were from the *Cecile Essentials of Medicine* (2010), which is also the reference textbook for the main study. In text selection, the texts were selected at random from all samples, and to have a balanced corpus, each subcorpus contained at least 80,000 running words (see Table 4.2).

Table 4. 2: Subcorpora and running words of the medical academic corpus of the pilot study

Sub-corpus	**Psychiatry**	**Obstetrics & Gynecology**	**Pediatrics**	**Internal medicine**	**Surgery**
Running words	80,888	80, 988	80,618	80,664	80,220

The corpus was analyzed by the Compleat Lexical Tutor (CLT) version 8, and prepared for word selection to form the word list. The words for inclusion in the word list were selected through the unit of word families, i.e. a stem and its inflected and derivational forms, as well as all closely related affixed forms.

According to Coxhead (2000), studying the data of the Brown Corpus shows that "a corpus of around 3.5 million words would be needed to identify 100 occurrences of a word family" (p. 217). That is, the corpus of this pilot study which contains around 400,000 running words would be needed to identify at least 11 occurrences of a word family to meet the criterion of frequency for inclusion in the word list.

Overall, three key principles guided me in selecting the words for inclusion in the MAWLOT. These principles were specialised occurrence, range, and frequency.

1- Specialised occurrence: The words included in the MAWLOT had to be outside the GSL (West, 1953) and the AWL (Coxhead, 2000).

2- Range: The words included in the MAWLOT had to occur at least in 3 or more of the 5 subcorpora in the corpus.

3- Frequency: Members of a word family had to occur at least 11 times in the corpus. They had also to occur at least 2 times or more in each subcorpus.

The rationale for excluding words in the GSL (West, 1953) and the AWL (Coxhead, 2000) from the MAWLOT is to indicate that the list is specifically for medical students. According to Coxhead (2000), an academic word list (such as the MAWLOT) should contain the words that appear across the various subject areas covered by the texts contained within the corpus (p. 216). This methodological approach, i.e. excluding the GSL and the AWL to develop a specialised word list, has been already applied by some researchers. For example, Wang et al. (2008) excluded the GSL and the AWL's words in developing the medical academic word list (MAWL). Furthermore, Fraser in 2007 compiled the Pharmacology Word List (PWL) based on a corpus of pharmacy research articles, along with the GSL and the AWL. The resulting lexical coverage was 88% of the Fraser's corpus. Again, in 2009, Fraser developed a PWL consisting of only core

pharmacological words (2,000 words) by excluding the GSL and the AWL. The resulting lexical coverage was higher (89%) than the first PWL in his corpus.

Regarding the present study, it is obvious that the GSL (West, 1953), with 2000 most frequently occurring words of English "with the needs of ESL/EFL learners in mind" (Coxhead, 2000, p. 213), and the AWL (Coxhead, 2000) with 570 word families "for students of commerce, art, law, and science" (Coxhead, 2000, p. 223) do not contain predominantly medical academic words for medical students, as the target learners of the present study.

To develop the word list, I used the section of *VocabProfiler v.4* from the Compleat Lexical Tutor (CLT) software program. The program is user-friendly and processes the data well and fast. Nevertheless, the CLT is an online program and the speed of data processing depends on the speed of the Internet availability.

Forty four word families met the criteria for inclusion in the primary word list. To form the final list, the primary list was justified by removing 17 words (see Appendix C for the removed words from the list) to study only the medical academic words necessary for academic reading and found often in medical academic textbooks. Some of the most frequent words in the pilot study were *mortal, clinic* and *infect*. The words of *random*, *figure*, and *America* were among the removed words. Obviously, the number of words which met the criteria for inclusion in the word list has been obtained from a small corpus of the pilot study, and is not enough to increase the vocabulary knowledge of medical students in reading medical academic textbooks at the general physician stage.

In addition to identification of words, finding particular patterns of the words included in the list was needed. The AntConc software version 3.4.4 was used to find

the patterns of the words included in the MAWLOT by analyzing each subcorpus. The software is a freeware, offline, multi-platform tool for carrying out corpus linguistics research and data-driven learning.

The method to investigate the particular patterns of the words included in the MAWLOT is extensively explained in section 4.8.2. To investigate the patterns of the words in the word list of the pilot study, a number of concordance lines (20 lines), which had been sorted into alphabetical order, were selected at random from all retrieved concordance lines. These lines contained the examined word to investigate the patterns (Hunston & Francis, 2000). The concordance lines were sorted to the right to assist in pattern identification.

Three words, i.e. *disorder*, *therapy*, and *patient* were selected from the word list of the pilot study (see Appendix B) to consider their particular patterns in the corpus. The words were among the top frequent words of the word list and revealed the same frequency and range (F=36, R=5, %= 0.009). The subcorpora of the study revealed the following significant patterns:

1. Disorder

a. internal medicine

The subcorpus of internal medicine reveals the pattern **adj N**, which means that the noun *disorder* is preceded by an adjective:

adj N

9 ant urinary retention, patients with a neurologic disorder such as multiple sclerosis or spinal cord

b. obstetrics and gynecology

The pattern **a/an adj N** is revealed by the subcorpus of obstetrics and gynecology, which means that **N** is preceded by an adjective and the articles *a* or *an*:

a/an adj N

1 (see Chap. 34, Preeclampsia as a Two-Stage Disorder). Modifications of spiral arteries are ca

2 biosynthesis. This deficiency is an X-linked disorder, and all affected fetuses are male. Its

c. pediatrics

The subcorpus of pediatrics reveals the pattern **N is pp prep**, which means that the noun is followed by the specific verb *is*, a past participle, and a preposition. The other patterns are **adj N**, **adj N**, **a/an adj N**, **prep adj N, N is characterized by**, and **factitious N or malingering**:

adj N

20 for by a mood, anxiety, or psychotic disorder and does not meet criteria for dyspareuni

23 is not considered to be a mental disorder and is included here to facilitate differ

a/an adj N

37 the suspicion that they have a genetic disorder are called probands; individuals who come

N is pp prep

16 subsequent development of panic disorder. Panic disorder is described in all racial groups. Lifeti

prep adj N

21 fatigue usually none very good for hepatic disorder; if myopathy present, it tends to be

N is characterized by

11 o decrease morbidity is important. Acute stress disorder is characterized by the same signs and

12 ormation is unclear or contradictory. Cyclothymic disorder is characterized by 2 years or more (1 ye

factitious N or malingering

12 not intentionally produced or feigned (factitious disorder or malingering) after appropriate investi

17 symptoms. Somatization disorder is not a factitious disorder or malingering. These syndromes commonly

d. psychiatry

The selected concordance lines reveal the pattern **prep adj N**, which means that the noun is followed by the verb *is*, a past participle, and a preposition. The other patterns are **adj N**, and **n N or n**. The pattern **n N or n** also reveals that the noun is preceded by a noun, and it is followed by the specific preposition *or* and another noun:

adj N

18 mage to the left hemisphere produces intellectual disorder and loss of the narrative aspect of

30 well-being presumes the absence of mental disorder defined in the text revision of the

prep adj N

48 concert with its salutary effects on bipolar disorder. Likely, ongoing investigations will imply

n N or n

25 with asthma meet the criteria for panic disorder or agoraphobia. The fear of dyspnea can

40 when the person already has a seizure disorder or brain lesion. Sedation Blockade of h

e. surgery

prep adj N

21 Historically, the two broad categories of mental disorder are psychosis and neurosis. Psychosis is

48 concert with its salutary effects on bipolar disorder. Likely, ongoing investigations will imply

2. Patient

a. internal medicine

The subcorpus reveals the pattern **the N and family**, which means that the noun is preceded by the article *the*, and followed by the conjunction *and*, and the specific noun *family*. The other patterns revealed by the subcorpus are **a N with adj**, **N with adj**, **N with n**, **prep the N and physician**, **prep the N is**, and **the N is –ing**:

the N and family

3 of illness, ensuring safe passage for the patient and family and guiding them through an

5 that death is likely without leaving the patient and family feeling hopeless or abandoned:

a N with adj

5 serial test results can be misleading. Suppose a patient with chronic obstructive pulmonary disease

8 60 to 100 days because of the assumption that a patient with cutaneous anthrax as a result

N with ad

7 (Chapters 12 and 13) to the treatment of the patient with acute myocardial infarction and

9 test results can be misleading. Suppose a patient with chronic obstructive pulmonary disease

N with n

2 categories. Positive symptoms are symptoms that a patient with schizophrenia has that are added to

prep the N and physician

19 ; it should then be reviewed by the patient and physician and periodically updated.

22 measure of trust. The interaction between the patient and physician represents not only a scient

prep the N is

5 court declared that obtaining prior consent of the patient is not only the usual practice

the N is -ing

1 . Re-evaluate long-time drug use because the patient is changing. Review over-the-counter

b. obstetrics and gynecology

The lines reveal the pattern **the N and family**, and **prep the N and physician.** The pattern **prep the N and physician** means that the noun is followed by the conjunction *and*, and the specific noun *Physician*:

the N and family

10 remaining time while striving to reassure the patient and family that everything reasonable

prep the N and physician

19 ; it should then be reviewed by the patient and physician and periodically updated.

22 measure of trust. The interaction between the patient and physician represents not only a

c. pediatrics

The subcorpus of pediatrics reveals that the noun is preceded by the article *the*, and followed by the specific verb *is* in the pattern **the N is**. The other patterns are the **N is –ing**, and **the N and family**:

the N and family

2 the primary diagnosis and by helping the patient and family deal with the stresses and

5 resources, and evaluate the responses of the patient and family members to the treatment program

the N is

2 need to maintain the sick role. The patient is convinced that the symptoms are unrelat

the N is -ing

10 of the treatment and ensure that the patient is receiving therapeutic amounts of the me

11 understanding of the clinical situation, how the patient is responding, and the factors that may

d. psychiatry

The subcorpus reveals that the noun is followed by the specific verb *is*, and an **–ing** form in the pattern **N is –ing**. The other patter are **N with n**, **N doctor relationship,** and **the N is**:

the N is

10 preoccupations, the doctor may surmise that the patient is speaking about trivial or insignificant

14 doctor has not told the truth, the patient is unlikely to accept or believe truthful

the N is -ing

5 's expression of concern for what the patient is experiencing. The psychiatrist Harry St

13 or she has correctly understood what the patient is trying to say and to let

N with n

7 above such emotions and handle a difficult patient with equanimity, the interpersonal relation

13 to organize the rambling associations of a patient with schizophrenia. At the neurochemical l

N doctor relationship

6 . In some medical specialties in which the patient–doctor relationship is not particularly in

11 occurs in the course of a particular patient–doctor relationship, their care will be on

e. surgery

The subcorpus reveals the pattern **a N with adj**, which means that the noun is preceded by the article *a*, and followed by the specific preposition *with* and an adjective:

a N with adj

11 usually occurs in the setting of a patient with endemic goiter. Increased thyroid hor

3. Therapy

a. internal medicine

The subcorpus of internal medicine reveals the particular patterns **adj N**, **n N prep**, and **adj N and**:

adj N

3 gnition and avoidance of further insult. Physical therapy and medications may hasten recovery but ca

5 ntions, such as reassurance, cognitive-behavioural therapy, and placebo effects associated with drug

n N prep

15 nges in clinical practice, such as recanalization therapy for acute myocardial infarction (Chapter

adj N and

1 gnition and avoidance of further insult. Physical therapy and medications may hasten recovery but ca

3 ntions, such as reassurance, cognitive-behavioural therapy, and placebo effects associated with drug

b. obstetrics and gynecology

There is no discerning pattern similar as the above ones.

c. pediatrics

adj N

3 intenance phase is more effective than behavioural therapy alone. Maintenance medications, including

9 ng cognitive and behavioural therapies, supportive therapy, and group therapy, have been used. Group

adj N and

5 disorders. The individual components of cognitive therapy and behavioural therapy can be used alone

d. psychiatry

adj N

13 h less psychotherapeutic training than is marital therapy. Marital therapy places greater emphasis o

e. surgery

The subcorpus of surgery reveals the pattern **adj N for metastatic disease**, which means that the noun is followed by the specific preposition *for*, the specific adjective *metastati*c, and the specific noun *disease*. Other patterns are **adj N for metastatic disease, adj N**, **n N prep**, **adj N and**, **radiation N is**, and **radiation N and n**:

adj N for metastatic disease

18 (AIs) Anastrozole, letrozole, exemestane Adjuvant therapy for metastatic disease pure antiestrogens

adj N

1 therefore differ in their response to biologic therapy. A chimeric form of the monoclonal antibody

4 and thus emphasize the complexity of endocrine therapy. Adjuvant Hormonal Therapy for Operable Br

n N prep

2 The combination of chemotherapy and radiation therapy for advanced-stage disease (stages III and

adj N and

8 clinical trials are introducing SAIs into adjuvant therapy and comparing these agents with tamoxifen,

33 for systemic treatment (chemotherapy or endocrine therapy) and to predict the response of patients

radiation N is

1 ignancies is en bloc surgical excision. Radiation therapy is administered postoperatively for high-g

13 cal recurrence after resection. Because radiation therapy is of benefit in the treatment of

radiation N and n

13 egral role. Combining chemotherapy with radiation therapy and surgery for treatment of advanced sino

15 Ductal Carcinoma In Situ: Impact of radiation therapy and Tamoxifen local recurrence rate (%)

Table 4.3 shows the above revealed patterns in sum.

Table 4.3: The most frequent collocational patterns of the pilot study

No.	Word	Internal medicine	Obstetrics & Gynecology	Pediatrics	Psychiatry	surgery
1	**DISORDER**					
	adj N	3	-	10	15	-
	a/an adj N	-	2	1	-	-
	N is characterized by	-	-	6	-	-
	factitious N or malingering	-	-	3	-	-
	n N or n	-	-	-	5	-
	prep adj N	-	-	4	4	5
2	**PATIENT**					
	the N and family	11	10	7	-	-
	N with adj	5	-	-	-	-
	a N with adj	3	-	-	-	5
	N with n	3	-	1	8	-
	prep the N and physician	5	4	-	-	-
	the N is	-	-	11	10	-
	prep the N is	5	-	-	-	-
	the N is -ing	1	-	2	5	-
	N doctor relationship	-	-	-	8	-
3	**THERAPY**					
	adj N	9	-	13	2	7
	adj N and	2	-	4	-	1
	adj N for metastatic disease	-	-	-	-	4
	radiation N is	-	-	-	-	6
	radiation N and n	-	-	-	-	5

4.3.2 The benefits of the pilot study to conduct the main research

By doing the pilot study, it helped me to assess the corpus of the study (the Medical Academic Corpus) with regard to size and the representativeness. Meanwhile, the instruments of the study were examined for their adequacy to conduct the main study. Examining the sample corpus (the corpus of the pilot study) through the instruments, enabled me to estimate the appropriate size of the corpus for the main study. Also, it confirmed that the instruments were adequate to apply in the main study to fulfil the objectives of the study. That is, through the pilot study a small sample of words which met the criteria for selection of the study was identified to include in the specialised word list. Seventeen words were removed from the list to justify the list as the medical academic word list. Twenty seven words remained to include in the final list of the pilot study (see Appendix B).

Two points were cleared by the pilot study. First, the corpus is small in its size to present an appropriate medical academic word list (27 words however useful but are not enough to meet the medical students' needs). Second, the extracted list of the study has to be modified to include just medical academic words. Therefore, the size of the corpus for the main study was increased to 3.5 million words (similar to the study of Coxhead (2000) in developing the AWL) to develop the MAWLOT. Also, the obtained list of the pilot study was finalised by removing 17 words, which however they met the criteria for inclusion in the MAWLOT, but they were not medical academic words to include in the medical academic word list based on textbooks. The criteria to select the words for inclusion in the word list were confirmed as well.

In addition, analysing the subcorpora of the pilot study revealed some particular patterns for the words included in the list. It also appeared that there are differences in the words' behaviours in each subcorpora, which have to be considered in the main

study. For example, the word *disorder* reveals the particular patterns **N is pp prep**, **N is characterized by**, and **factitious N or malingering** in the subcorpus of pediatrics. In the subcorpus of psychiatry, the noun *disorder* reveals the patterns **N is pp prep**, and **n N or n**. The discussion on differences between the behaviours of the MAWLOT's words is considered in chapter 6. Finally, doing the pilot study revealed that the instruments of the study, i.e. the Compleate Lexical Tutor (CLT) version 8, and the ActConc software version 3.4.4 are suitable to be employed by the study to analyze the data.

4.4 Data collection

The main goal of the current study was to develop a medical academic word list of textbooks. The Medical Academic Corpus was compiled with 3.5 million running words. The criteria to select the words for inclusion in the word list were recognized. To retrieve the corpus information to develop the word list, the Compleat Lexical Tutor (CLT) was selected as the instrument to analyze the data. Analyzing the data will answer the research questions of the study.

4.5 The method of the Medical Academic Corpus compilation

Based on the guidelines suggested by the literature review and sample framing, the Medical Academic Corpus was compiled for the study as the main resource for selecting words. Next, the words were selected to develop the Medical Academic Word List of Textbooks (MAWLOT). In this topic, I describe the corpus compilation of the medical textbooks.

4.5.1 The Medical Academic Corpus

This corpus was specifically compiled from academic texts in the field of medicine based on the sample framing. It was used as the main linguistic resource for word selection to be included in the medical academic word list of textbooks. The main

considerations in compiling the corpus were text selection and corpus size in order to make the corpus representative and balanced.

Developing the corpus requires to collect all texts in the electronic form with full text and bibliographic information, which can be done by all computer programs (e.g. the *Word* program). This was then changed to the *.txt.* format. Changing to the *.txt* format standardized the selected textbooks to include in the corpus by removing all photos, tables and figures which were not readable by computer. The normalization of words was done by the program software, i.e. Compleat Lexical Tutor (CLT) version 8, as the computer software for processing the data included in the Medical Academic Corpus. CLT is able to read all inflections and derivations of a word. Next, the words were counted by the software. Counting the running words by the software program enabled me to balance the subdisciplines of the corpus based on the length of texts. According to Coxhead (2000), having an organized corpus divided into coherent (balanced) sections of equal size is essential to measure the criterion of range of the academic vocabulary across the different disciplines of the corpus.

4.5.2 Text selection

Text selection was based on the population and sampling frames of the study. According to Aston (2002), text topics should be clustered in particular areas. Therefore, in the present study only texts with topics related to the major fields of medicine, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery were selected in order to reduce dispersion of medical subjects. For example, I included cardiology texts since they are related to internal medicine in the corpus, but I excluded dentistry, which is not considered a subfield of medicine (see Table 4.1) in training general physician students (GPs) at authentic medical universities and Iran as well.

Apart from the text topics, only written textbooks (no other genres such as research articles) were used as the text type in the present study to compile the corpus. Details about the selected textbooks are illustrated in Table 4.4.

Table 4.4: The Sources to compile the Medical Academic Corpus with tokens and types

Subcorpus	**types**	**tokens**
Internal medicine		
Cecile Essentials of Medicine	22250	706664
Obstetrics & Gynecology		
Williams Obstetrics	13767	484273
Current diagnosis and treatment: Obstetrics and gynecology	7764	154712
Comprehensive Gynecology	5665	62003
Pediatrics		
Nelson essentials of pediatrics	12352	653592
Pediatrics: A Competency-Based Companion	6811	47026
Psychiatry		
Kaplan & Sadock's synopsis of psychiatry: behavioural Sciences/Clinical Psychiatry	27979	702,888
Surgery		
Sabiston textbook of surgery	12082	531785
Current Surgical Therapy	10600	170435

4.5.3 Corpus size

In the Medical Academic Corpus, the length of each selected text ranged equally around 700,000 running words to compile the main corpus consisting of about 3.5 million running words (comparable in size to the academic corpus developed by Coxhead (2000) to create the AWL), with 2,659 different types of words. According to

O'Keeffe et al. (2007, cited in Aluthman, 2017), in order to make a reliable generalisation, a minimum size for a corpus should not be less than 1,000,000 words.

In addition, dividing a corpus into subcorpora is a good strategy because a small size of each subcorpus is more manageable and available to be selected according to the focus of the study (Aston, 2002). Therefore, all running words in the Medical Academic Corpus were selected and kept in five subcorpora, namely: internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery. Each subcorpus included texts related to its subject area in medicine. Table 4.5 shows the subcorpora and running words included in each subcorpus.

Table 4.5: Subcorpora and running words of the medical academic corpus

Subcorpus	**Internal medicine**	**Obstetrics & Gynecology**	**Pediatrics**	**Psychiatry**	**Surgery**
Running words	706,664	700, 988	700,618	702,888	702,220

Table 4.5 introduces the sources to develop the corpus of the study with their types and tokens in detail.

4.6 Instruments

The Medical Academic Corpus was analyzed by the Compleat Lexical Tutor (CLT) version 8, an online software program (*http://www.lextutor.ca*) that contains a concordancer, a vocabulary profiler, an exercise maker, interactive exercises, and much more (see Figures 4.1, 4.2, and 4.3 for illustrations of the CLT). The software package is used extensively to perform lexical text analysis to measure the proportions of low-frequency and high-frequency vocabulary used by speakers of varying abilities. The software of the Compleat Lexical Tutor, designed by Tom Cobb et al. (2005), is the

adapted form of the *Range* (Healty & Nation, 1996), the software used to develop the AWL (Coxhead, 2000). The section of the *VocabProfiler* can be used with any text to classify the words into four categories by frequency, including the 2,000 most frequent words of West's (1953) General Service List of English Words (GSL) (first most frequent 1,000 words in English, and the second most frequent thousand words, i.e. words 1,001 to 2,000), academic word list (AWL), and the remaining items which are not found in the GSL and in the AWL. The CLT counts the frequency of words in up to 25 files at a time and records the number of files in which each word occurs (range) and the frequency of occurrence of the words in total and in each file. The *VocabProfiler* has two ways of conveying the data: 1) input method A, in which a user can type or paste short to medium-sized text around 35,000 words into a window, and 2) input method B, in which the user can upload larger files up to 250,000 words.

Furthermore, the software is divided into three main sections: *Tutorial, Teachers*, and *Research*. The *Tutorial* provides self-access learning opportunities for learners. In the second section, *Teachers*, instructors can input their own customized vocabulary practice materials for their students. The third section, *Research*, is the most relevant to this study. This section is divided into eight sub-sections: 'Text tools' or tools for processing text, such as frequency list-makers, corpus builders (for those who wish to create their own corpora), a sentence extractor that removes end punctuation from sentences in texts, and HTML-tag strippers (useful for rendering HTML files into text files for further analysis). *'Range'* is another sub-section which enables the researcher to locate words and phrases in different corpora; 'Vocab Profile - the Web Vocabulary Profiler'; 'Concordance tools' with links to English and French concordance programs; 'Phrase Extractor' for analysing text for repeated phrases of 2-5 words to enable researchers to carry out the collocational analysis; and the 'RT Expt. Builder (Reaction Time Instrument Builder), a tool used to test the speed of learners' word recognition.

‘Vocab Stats’ is a collection of programs for researchers who wish to generate different types of statistical information about patterns of vocabulary in texts. It links to *Richard Lowry's Concepts and Applications of Inferential Statistics* at Vassar (Cobb & Horst, 2004). And finally, the 'Why & wherefore/Research Base' contains links to a number of Cobb's publications and research, as well as other studies and websites dealing with the topic of 'list and frequency-based vocabulary learning'.

With its focus on academic vocabulary, the software program and its many features have great potential for those working with English for Academic Purposes (EAP) (Cobb & Horst, 2004). It also obviously has great potential for those interested in profiling the medical words and typical phrases of the various disciplines in medicine, and for comparing those profiles in terms of frequencies of words and word families, collocational tendencies, range and other aspects.

CLT has been previously used in applied linguistics in the study of Farjami (2014) to explore the frequency of words in the abstracts of applied linguistics journal articles. The resulting list included a list of the most frequent content words, lists of frequent words and abbreviations not found in the British National Corpus. It has also to be said that, like any other software, the CLT has its own weaknesses points. For example, as Cobb (2004) states, one of the minimum system requirements through using the CLT is to have an Internet line with a modem of 56K, which reveals the dependency of the software to the online system. Therefore, working with the software is difficult through the low speed of the online system. I faced with this problem through using the CLT in the present study as well.

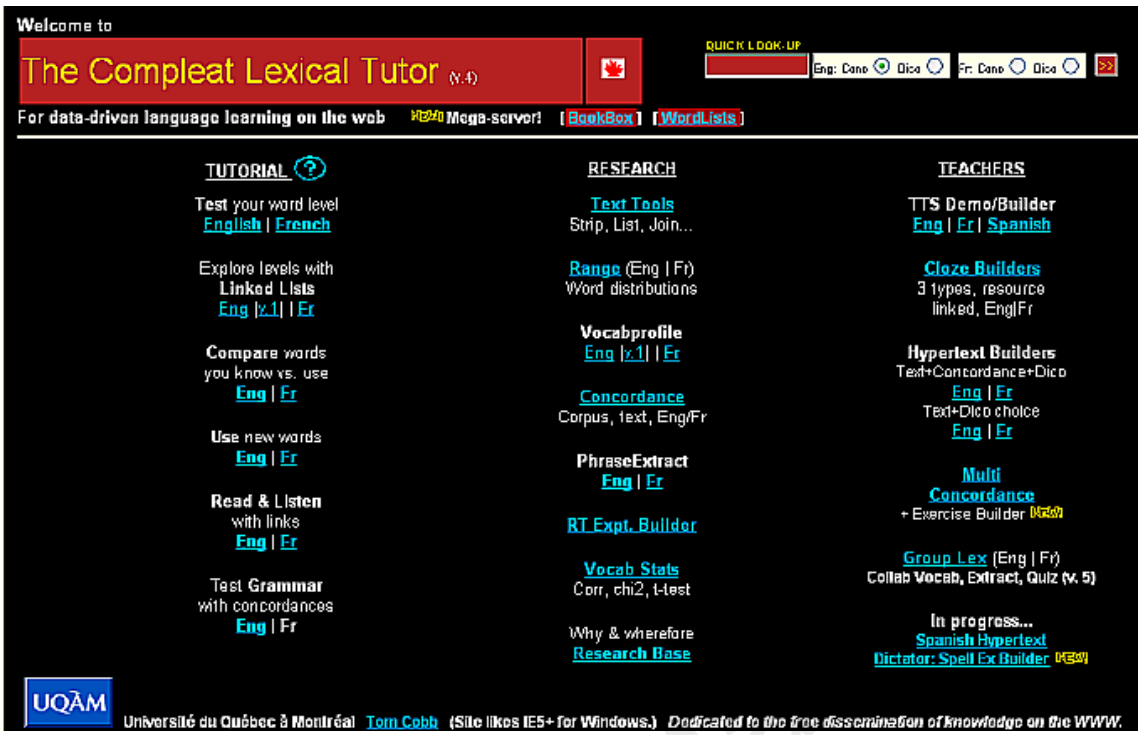


Figure 4.1: The Compleat Lexical Tutor, site entry web page (Cobb et al., 2005)

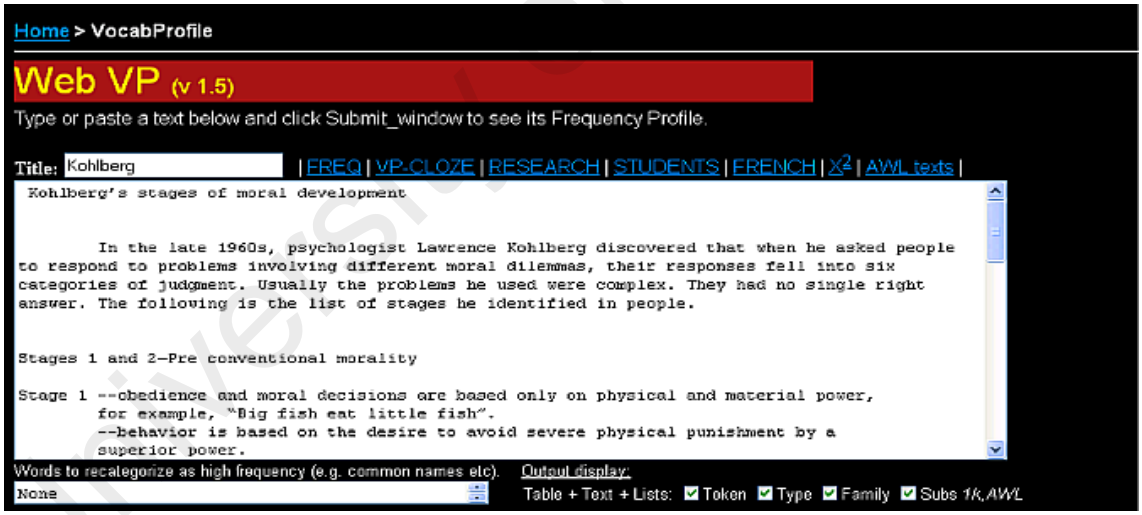
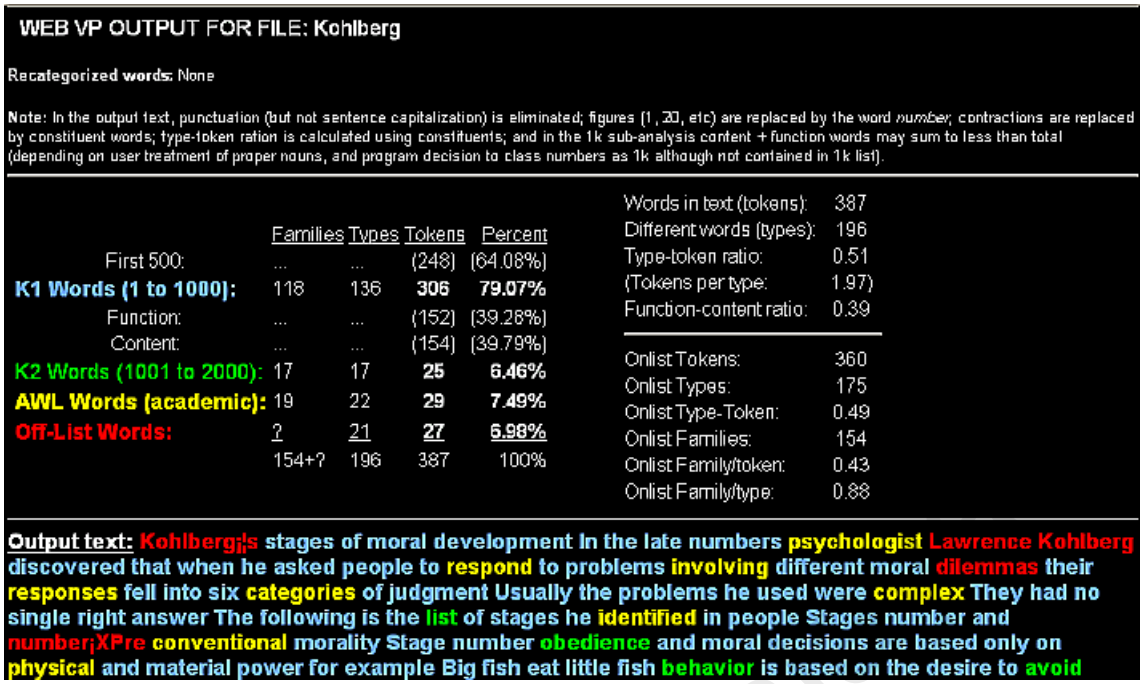


Figure 4.2: Web Vocabulary Profiler, input (Cobb et al., 2005)



WEB VP OUTPUT FOR FILE: Kohlberg

Recategorized words: None

Note: In the output text, punctuation (but not sentence capitalization) is eliminated; figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words; type-token ration is calculated using constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns, and program decision to class numbers as 1k although not contained in 1k list).

	Families	Types	Tokens	Percent
First 500:	...	...	(248)	(64.08%)
K1 Words (1 to 1000):	118	136	**306**	**79.07%**
Function:	...	...	(152)	(39.28%)
Content:	...	...	(154)	(39.79%)
K2 Words (1001 to 2000):	17	17	**25**	**6.46%**
AWL Words (academic):	19	22	**29**	**7.49%**
Off-List Words:	?	21	**27**	**6.98%**
	154+?	196	387	100%

Words in text (tokens):	387
Different words (types):	196
Type-token ratio:	0.51
(Tokens per type:	1.97)
Function-content ratio:	0.39
Onlist Tokens:	360
Onlist Types:	175
Onlist Type-Token:	0.49
Onlist Families:	154
Onlist Family/token:	0.43
Onlist Family/type:	0.88

Output text: Kohlberg¡s stages of moral development In the late numbers psychologist Lawrence Kohlberg discovered that when he asked people to respond to problems involving different moral dilemmas their responses fell into six categories of judgment Usually the problems he used were complex They had no single right answer The following is the list of stages he identified in people Stages number and number¡XPre conventional morality Stage number obedience and moral decisions are based only on physical and material power for example Big fish eat little fish behavior is based on the desire to avoid

Figure 4.3: Web Vocabulary Profiler, output (Cobb et al., 2005)

Another instrument used in the research was the AntConc's software version 3.3.5. This version was released in July 2012 (Anthony, 2012). AntConc is a freeware application which runs on both Windows and Linux systems. The software program is available at the URL address *http://www.laurenceanthony.net/software/antconc/*. As can be found in the software, AntConc contains seven tools that can be accessed either by clicking on their 'tabs' in the tool window, or using the function keys F1 to F7. These tools are Concordance Tool, which shows search results in a 'KWIC' (KeyWord In Context) format to allow researchers to see how words and phrases are commonly used in a corpus of texts. Concordance Plot Tool, which shows search results plotted as a 'barcode' format to allow researchers to see the position where search results appear in target texts. File View Tool, which shows the text of individual files to allow researchers to investigate in more detail. The Clusters Tool, which shows clusters based on the search condition. Basically, it summarises the results generated in the Concordance Tool or Concordance Plot Tool. The N-Grams Tool, which scans the entire corpus for 'N' (e.g. 1 word, 2 words, etc.) length clusters to allow researchers to

find common expressions in a corpus. This tool is used extensively by researchers in corpus- based studies. Figure 4.4 shows the Clusters/N-Grams tool of AntConc.

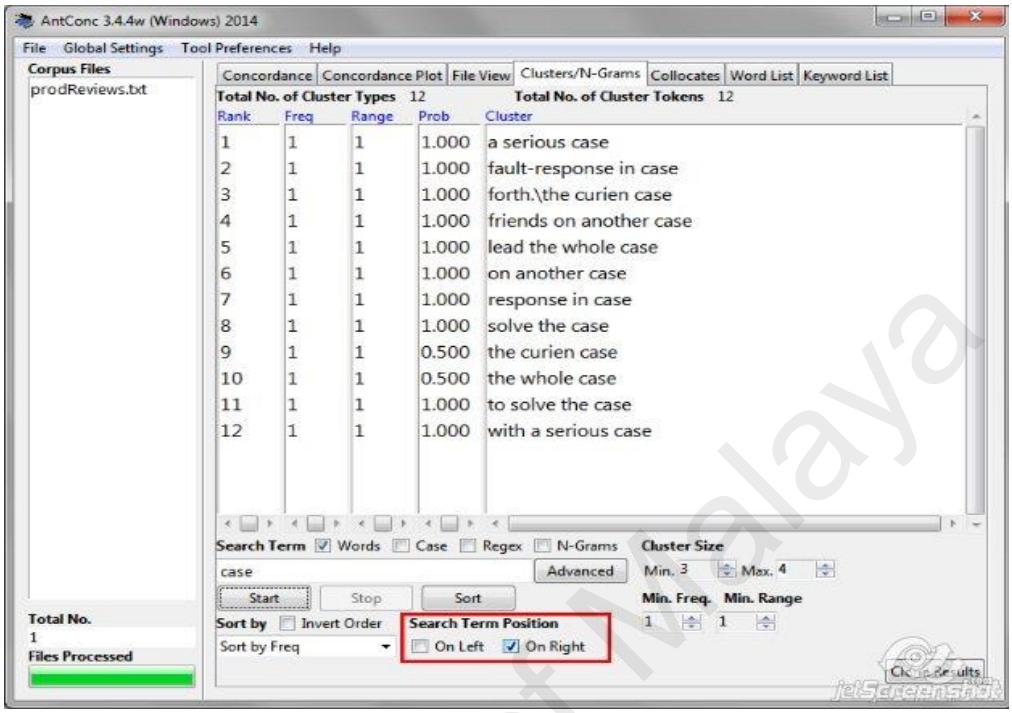


Figure 4.4: AntConc Clusters/N-Grams tool (Anthony, 2012)

The Collocates tool, which shows the collocations of a search term to allow researchers to investigate non-sequential patterns in language. The Word List, which counts all the words in the corpus and presents them in an ordered list to allow researchers to quickly find which words are the most frequent in a corpus. The Keyword List, which shows the which words are unusually frequent (or infrequent) in the corpus in comparison with the words in a reference corpus. This allows researchers to identify characteristic words in the corpus, for example, as part of a genre or ESP study. Figure 4.5 shows the concordance tool of the program.

Figure 4.5: AntConc Concordance tool (Anthony, 2012)

AntConc has a freeware license, and it is used extensively by the researchers in corpus-based studies (e.g. Farjami, 2016; Noguchi, 2004), and has been described as extremely effective in a classroom context by Noguchi (2004). The only problem I faced with through applying the AntConc in the study was the small number of items in output. For example, the total number of types reported by the software for the Medical Academic Corpus with 3,513,376 tokens was 60699 words. While, for the subcorpus of internal medicine with 706,664 tokens the number of words' types was 22250, for the subcorpus of obstetrics and gynecology the tokens were 700,988 with 27196 different types of the words. In addition, for the subcorpus of pediatrics with 700,618 tokens the reported types of the words was 19163, for the subcorpus of psychiatry the number of tokens was 702,888 with 27979 types, and for the subcorpus of surgery with 702,220 tokens the reported types of the subcorpus was 22682. As Farjami (2014) states, in using the AntConc it is a rule that the longer the input, "the smaller the number of items in the output" (p. 57). It is due to the matter of the word frequencies included in the

input (here, corpus) that, "the more frequent the items, the smaller the number of the items with similar frequencies" (p. 58).

4.7 Word selection

To develop the Medical Academic Word List of Textbooks (MAWLOT), all words for inclusion in the word list were selected through the unit of word families — defined as a stem and its inflected and derived forms, as well as all closely related affixed forms plus the most frequent stem, perceived transparency, and regular and productive prefixes and suffixes (Bauer & Nation, 1993). In addition, the Medical Academic Corpus was organized to allow the principle of range to consider the occurrence of particular words to be examined. Similar to studies of Coxhead (2000) and Wang et al. (2008), the criterion of frequency was deemed secondary to range to avoid bias in favour of longer texts and topic-related words. This means that the words that occurred with a high frequency but did not occur in at least 3 or more subdisciplines out of 5 were removed.

Three key principles guided us to select the words for inclusion in the MAWLOT. These principles were specialised occurrence, range, and frequency.

1. Specialised occurrence: The word families included in the MAWLOT had to be outside the GSL (West, 1953) and the AWL (Coxhead, 2000) (see pp. 62-63).

2. Range: The word families included in the MAWLOT had to occur in at least 3 or more of the 5 subcorpora in the Medical Academic Corpus.

This measure (range) serves to ensure that a selected lemma occurs in a wide range of subcorpora. In development of the AWL and the MAWL, Coxhead (2000) and Wang et al. (2008) specified that the words should appear in at least half of the subcorpora.

3. Frequency: Members of a word family had to occur at least 100 times in the Medical Academic Corpus. I made this decision because "studying the data of the Brown Corpus (Francis & Kucera, 1979) shows that a corpus of around 3.5 million words would be needed to identify 100 occurrences of a word family" (Coxhead, 2000, p. 217). Coxhead (2000) and Wang et al. (2008) used the same threshold in developing the AWL and the MAWL, the lists with which the MAWLOT is compared (explained in section 5.3.3). Meanwhile, the words had to be occurred at least 20 times in each of the five subcorpora of the Medical Academic Corpus for inclusion in the word list. The words would be also statistically calculated to measure the percentage of the words' occurrences in the corpus.

4.8 Data analysis

The Medical Academic Corpus of the present study was analyzed through the Compleat Lexical Tutor (CLT) and AntConc software. An initial word list, the collocations of the most frequent words included in the list, and the coverage of each subcorpus by the word list were analyzed to find the adequate response for the research questions proposed in the study (see p. 7). The first three questions deal with descriptive phase of developing the Medical Academic Word List of Textbooks (MAWLOT), and explain how to construct the word list and the collocations of the words of the MAWLOT. The last three questions describes the evaluation phase of the study by considering the coverage of the MAWLOT of vocabulary of the Medical Academic Corpus and comparison this coverage with the coverage of the AWL (Coxhead, 2000) and the MAWL (2008).

4.8.1 Data analysis for Research Question 1

The first research question attempted to recognise the MAWLOT's lexical items in terms of frequency and range occurred in the corpus of medical textbooks beyond the

AWL (Coxhead, 2000), and the GSL (West, 1953). To answer the first research question, I used the Web VP Classic v.4 from the section of the 'VocabProfilers' (VP) of the Compleat Lexical Tutor (CLT). The VP is available at the URL address http://www.lextutor.ca/vp/, and is able to break texts down by word frequencies in the language at large as opposed to in the text itself. According to information included in the CLT's site version 8 (2014), the English Vocabprofilers on the software program are based on Laufer and Nation's Lexical Frequency Profiler (1995) 4-way sorter, which divide the words of texts into the first and second thousand levels, academic words, and the remainder or 'off list'. The VP section is user-friendly and gives me the needed corpus information to form the primary word list.

The final word list was developed after modifying the primary list with 665 words. That is, however all 665 words met the criteria for inclusion in the MAWLOT, 206 words were removed from the list to study only those medical academic words that are necessary for academic reading and found often in medical academic textbooks (see Appendix E for the removed words and their frequency and range). For example, the words *Africa, America, edit, copyright, school, electronic, hour, junior, year,* etc. were removed from the primary list. The resulting word list formed 'the Medical Academic Word list of Textbooks' (MAWLOT) with 459 words outside the GSL (West, 1953) and the AWL (Coxhead, 2000), and with high frequency in the Medical Academic Corpus, which will be explained in Chapter 5.

A list of medical technical words was also retrieved to increase the learners' vocabulary knowledge in reading medical texts (see p. 99). Those words that are particularly unique and are used only in the discipline of medicine are counted as the medical technical words (see section 2.5.1.3). These words are discipline-specific words (domain-specific academic words or content- specific words) (Hiebert & Lubliner,

2008) and have just one meaning that understanding it is essential to build conceptual knowledge in the discipline they are using. For example, the words *tumors, anemia, hemoglobin, gonadotropin* are counted as technical words in medicine.

4.8.2 Data analysis for Research Question 2

Collocation analysis is one of the most extensively used methods in corpus linguistics today (Lehecka, 2015). Therefore, the second research question attempted to identify MAWLOT's most frequent collocations in the medical academic word list of textbooks to find the patterns of the words, and consequently to observe the words' behaviour in context. The purpose of this statistical calculations is to find word combinations with significantly more co-occurrences than what would be expected by chance, given the words' total frequencies in the data (Lehecka, 2015). Therefore, I can establish the most significant collocates of the MAWLOT's words based on the Medical Academic Corpus data. "If the observed number of co-occurrences of W1 and W2 is larger than what can be ascribed to chance, then W2 is a statistically significant collocate of W1" (Lehecka, 2015, p. 2).

Therefore, similar to the pilot study, after computational analysis and refinement of the data-driven list, the corpus was analyzed to find the most frequent and pedagogically relevant patterns to apply by EAP (EMP) teachers and medical students. Finding particular patterns of the most frequent words of the word list, i.e. Medical Academic Word List of Textbooks (MAWLOT), requires the application of a proper concordancer. Using the concordancer helps to search and obtain the needed information quickly. As said in the pilot study, ActConc software version 3.4.4 was used to find the patterns of the words included in the MAWLOT by analyzing each subcorpus. In addition, due to the Kjellmer's (1984) suggestion, in a corpus such as the Brown corpus with one million words, a string of words must happen more than once in

order to be counted as a collocation. Therefore, in the present study with 3. 5 million running words I select the collocations which have occurred at least 3 - 4 times in the corpus.

Finding the particular patterns requires to follow these steps:

1. Analyzing each subcorpus (5 subcorpora: internal medicine, obstetrics and gynecology, pediatrics, psychology, surgery) by the Word List part of the AntConc software.

2. Finding the most frequent patterns for the words of the MAWLOT's list (20 statistically top words) by using N-Grams part of the software. For example, analyzing the word *patients* **(PL-N)** in the subcorpus of internal medicine**,** as the top word included in the list of MAWLOT by N-Grams, shows that the most frequent part of speech which results immediately to the right of the **PL-N** are *with* **(PL-N with)** and *in* **(PL-N in)**. The symbol **PL-N** is in uppercase as we are focusing on a plural noun (Hunston & Francis, 2000).

3. Finally, according to Hunston and Francis, the procedure for investigating the patterns of a word involves selecting at random a number of concordance lines which have been sorted into alphabetical order. They selected 50 concordances from all retrieved concordance lines which contained the examined word to investigate the patterns (I also follow this procedure). Depending on the word, the concordance lines may be right-sorted or left-sorted. In the case of a verb, noun, and an adjective it is better to sort them to the right to reveal their complementation patterns (e.g. *the diagnosis of acute MI, in patients with a trichobezoar*). Therefore, the concordance lines of the study were sorted to the right because the MAWLOT's words are mostly nouns and adjectives.

In addition, the method of Francis et al. (1996; 1998) in representing the patterns included in the Collins COBUILD English dictionary (CCED), applied by Hunston and Francis (2000) in developing the COBUILD grammar patterns series, is used to represent the patterns of the study. The Collins COBUILD Grammar Patterns has two volumes: Collins COBUILD Grammar Patterns 1: Verbs (Francis et al., 1996), and Collins COBUILD Grammar Patterns 2: Nouns and adjectives (Francis et al., 1998). Therefore, "only the simplest and most superficial word-class labels" (Hunston & Francis, 2000, p. 45) are used to encode the grammar. It means that, **v** is used to encode the verb group, **n** as noun group, **adj** to encode adjective group, **-ing** to encode clause introduced by an '-ing' form, and **pl-n** to encode a plural noun. Besides, Hunston and Francis (2000) do not include *a* and *an* in representing the patterns when a verb, noun, and an adjective are followed by them. For example, the sample concordance line to represent the pattern **it V n** is "*It's blowing a gale"* (p. 56). Also, the sample concordance line to represent the pattern **V n** is *He played a trick* (p. 62). Furthermore, they have mostly selected just one concordance line to represent the pattern. I will also represent the patterns in 1-2 concordance lines of the examining word in the study.

Chapter 6 presents the collocation/phraseological patterns of the MAWLOT and the medical technical words, which are revealed most frequently in the Medical Academic Corpus of textbooks of the current study. The usefulness of teaching collocation/phraseological patterns will be also discussed.

4.8.3 Data analysis for Research Question 3

Question 3 aimed to find out if the words included in the MAWLOT occur with different frequencies in the subcorpora of medicine or not. To analyze the data, the descriptive data obtained from the Research Question 1 (i.e. the selected words for inclusion in the MAWLOT) were examined in each subcorpus in the section of *Text Lex Compare (TLC)* v.3, separately. TLC is available in the *Research* part of the CLT.

According to the Compleat Lexical Tutor (2005) site, the TLC is able to subtract one text from another, one list from another, and a list from a set of lists or texts (news articles, book chapters, etc.). TLC compares texts in 10 different ways according to the user's aim. Regarding the present study, a part of the program which can compare a text with the user list is used to compare the subcorpora of the study with the words included in the MAWLOT to find the differences in occurrences of words across the subcorpora. The software program has two methods to process the data: method A and method B. Method A is for small texts, and Method B is for uploading texts with a high volume. TLC is user-friendly. Also, it does not have any limitation on the text size to upload, but the speed of processing the data is quite a bit faster when the texts are smaller.

The output dedicates two important lexical items: Tokens Recycling Index (TRI) and Families Recycling Index (FRI). According to the program, token recycling index (TRI) is normally the most interesting measure of text comprehensibility, which achieves the division of the repeated tokens into the tokens in the new text. The different amounts of the TRI in different subcorpora show differences in occurrence of the words included in the MAWLOT across the subcorpora of the study.

The results of Research Question 3 can also lead us to find the answer for Research Question 4, which is the first research question of the evaluation phase.

4.8.4 Data analysis for Research Question 4

Data analysis for Research Question 3 shows differences in occurrences of the Medical Academic Word List of Textbooks (MAWLOT) across subcorpora of the Medical Academic Corpus. These amounts also show the coverage of each subcorpus by the MAWLOT, and consequently the total coverage of the corpus beyond the GSL (West, 1953) and the AWL (Coxhead, 2000) to answer the Research Question 4. This analysis has been done to check and establish the viability and representativeness of the

MAWLOT. If the coverage of the MAWLOT in an academic corpus is higher than the coverage of the AWL or the MAWL, then it would suggest that the MAWLOT contains a list of medical academic words rather than the AWL and the MAWL. To enable this analysis, I used the Medical Academic Corpus of the study as the medical corpus.

The results showed that the total coverage of the MAWLOT of vocabulary of the Medical Academic Corpus in the present study was 11.27%, higher that the coverage of the Academic Corpus by the AWL (10.0%), and close to the total coverage of the MAWL of vocabulary of the corpus in the study of Wang et al. (2008) (12.24%) based on the medical research articles (explained in section 5.3.1).

4.8.5 Data analysis for Research Question 5

A frequency-based word list of a particular corpus expects to cover the corpus well, but to really test the word list it has to be examined in similar texts of a different collection (Coxhead, 2000). To consider how representative is the MAWLOT of medical texts beyond the corpus of the study and whether the MAWLOT maintains the high coverage over medical academic texts of the corpus, I compiled a second corpus by using the same criteria and sources to select texts. The texts had not been included in the main corpus because the needed materials (running words) for the Medical Academic Corpus had already been completed. To have a balanced corpus I collected a corpus of 350,000 running words, i.e. about 70,000 tokens of each subcorpus (i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery), and called it the Second Corpus. Table 4.6 shows the details of the second medical academic corpus.

Table 4.6: Subdisciplines and running words of the second medical academic corpus

Subcorpus	**Psychiatry**	**Obstetrics & Gynecology**	**Pediatrics**	**Internal medicine**	**Surgery**
Running words	70,712	70,001	70,554	70,908	70,267

The results of the data analysis show how the word families included in the MAWLOT occur differently in the subcorpora of the second medical academic corpus with a higher coverage (see p. 105).

4.8.6 Data analysis for Research Question 6

A comparison of the obtained results of the above data analysis with the reported results in developing the AWL (Coxhead, 2000) and the MAWL (Wang et al., 2008) regarding tokens, types, and running words of the corpora in addition to coverage of each word list make answering Research Question 6 possible. Doing the analysis, I used the Medical Academic Corpus consisted of five subcorpora of internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery. I compared the total coverage with the coverage of the AWL and the MAWL to find out if the coverage of the MAWLOT in the Medical Academic Corpus is high enough to provide evidence that the MAWLOT contains academic words used more frequently in medicine rather than in the general academic English of Coxhead (2000) and the medical academic word list of Wang et al. (2008). If so, it would suggest that the MAWLOT contains a list of medical academic words rather than a list of general academic words (AWL) or medical words (MAWL), which are frequently found in medical research articles for specialists rather than students.

4.9 Summary

This chapter described the research methodology of the study. The type of corpora needed to compile the main corpus of the study was recognised. The population and samples were defined, first. Next, a pilot study was conducted in order to try out the samples and the instruments. The results were used to provide guidelines for the implementation of the main study. The instruments were recognised as suitable and user-friendly for conducting the main study. Regarding the data collection in the main

study, the Medical Academic Corpus of around 3,500,000 running words (about 700,000 tokens in each subcorpus) was compiled from academic texts in medicine. In data analyses, 665 target words were first selected for the primary list from high-frequency words that occurred in a wide range in the corpus. Next, the primary list was modified and the final wordlist was developed as the Medical Academic Word List of Textbooks (MAWLOT) including 459 words. The MAWLOT was used for answering Research Question 2-3 of the descriptive phase and Research Questions 4-6 of the evaluation phase. The following chapter and Chapter 6 will present the findings of the study.

CHAPTER 5: FINDINGS

5.1 Introduction

This chapter presents the findings from the study according to the main focus of the six proposed research questions, i.e. developing and evaluating the medical academic word list of textbooks. The first three questions centered upon describing the word list, while questions 4-6 dealt with evaluation of the word list. Therefore, the findings of descriptive and evaluation phases are presented respectively.

5.2 Findings of the descriptive phase (Research Questions 1- 3)

The first three research questions focused on developing a medical academic word list of textbooks (Research Question 1), collocation/phraseological patterns (Research Question 2), and considering the differences in word occurrences in five subdisciplines of medicine, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery (Research Question 3), as the descriptive phase of the study.

5.2.1 Occurrence of medical academic words

The findings of Research Question 1 first yielded a list of 665 word families. All the words met the criteria for inclusion in the MAWLOT, but the list needed to be modified to study only the medical academic words necessary for reading and which are found often in medical academic textbooks. Removing 206 words formed the Medical Academic Word List of Textbooks (MAWLOT) with 459 words (see Appendix D for the MAWLOT's list). Some of the most frequent word families in the MAWLOT are *patients* (F=8943, R=5), *disorder* (F=8457, R= 5), *therapy* (F= 3526, F= 5), *syndrome* (F=3519, R=5), and *fetal* (F=3498, R=5). Some of the least frequent words are *neuron* (F=100, R=4), *tubular* (F=100, R=5), *carcinoid* (F=100, R=4), and *oxidation* (F=100, R=5), (F=Frequency, R= Range).

The most frequent words among the word families included in the MAWLOT are generally the headwords of the word families. There are just a few headwords (15 headwords) that did not meet the criteria for inclusion in the MAWLOT. These headwords are: *bulimia* (headword: bulimic), *density* (headword: dense), *ejaculation* (headword: ejaculate), *excision* (headword: excise), *excretion* (headword: excrete), *facial* (headword: face), *fungal* (headword: fungus), *hypotension* (headword: hypotensive), *infarction* (headword: infract), *markedly* (headword: mark), *perforation* (headword: perforate), *schizophrenia* (headword: schizophrenic), *patients* (headword: patient), *signaling* (headword: signal), and *antisocial* (headword: social). Considering these words clears some differences in applying the words in medical textbooks. For example, the headword for the word family of *bulimic*, *bulimia* is *bulimic*, but the most frequent word selected for inclusion in the MAWLOT is *bulimia*, with a frequency of 128 in the Medical Academic Corpus of the study (the minimum frequency to include a word in the MAWLOT is F=100). According to the Stedman's Medical Dictionary (2012), *bulimic* refers to a person who suffers from *bulimia*, but *bulimia* (*bulimia nervosa*) is an eating disorder characterized by episodic binge eating, followed by depression and self-condemnation. Comparison of these two words shows that the word *bulimia* has occurred frequently in medical textbooks since it refers to a disease (not a person) with its own symptoms, signs, and treatments considerable by medical academic texts. Similarly, in the word family of *schizophrenic*, *schizophrenia*, the headword is *schizophrenic*, but the frequent word for inclusion in the MAWLOT is *schizophrenia* (F =160). Like *bulimic*, *schizophrenic* refers to a person as someone with a diagnosis of *schizophrenia* rather than as *schizophrenic*. *Schizophrenia* is a mental disorder that generally appears in late adolescence or early adulthood; however, it can emerge at any time in life. The disease may include delusion, loss of personality (flat affect), confusion, agitation, social withdrawal, psychosis, and bizarre behaviour.

Another example is the word family of *patient, patients*. The headword is *patient*, but the frequent word selected for the MAWLOT is *patients* with the highest frequency (F= 8943, R=5) among all the words included in the word list. That is, in spite of case reports or some medical articles, medical textbooks present general information about diseases and *patients*, and do not talk about a specific person.

Table 5.1 and Table 5.2 show the 50 top words and the 50 final words of the MAWLOT with their frequency and range.

Table 5.1: The fifty top words of the medical academic word list of textbooks with their frequency and range

Number	Word	Frequency	%	Range	Subdisciplines
1	patients (patient)	8943	0.25	5	T_1 T_2 T_3 T_4 T_5
2	disorder	8457	0.24	5	T_1 T_2 T_3 T_4 T_5
3	diagnosis	3765	0.10	5	T_1 T_2 T_3 T_4 T_5
4	therapy	3526	0.10	5	T_1 T_2 T_3 T_4 T_5
5	syndrome	3519	0.10	5	T_1 T_2 T_3 T_4 T_5
6	fetal	3498	0.09	5	T_1 T_2 T_3 T_4 T_5
7	cell	3428	0.09	5	T_1 T_2 T_3 T_4 T_5
8	clinical (clinic)	2107	0.06	5	T_1 T_2 T_3 T_4T_5
9	acute	2014	0.05	5	T_1T_2T_3T_4 T_5
10	infant	2000	0.05	5	T_1 T_2 T_3 T_4 T_5
11	drug	1979	0.05	5	T_1 T_2 T_3 T_4 T_5
12	cancer	1873	0.05	5	T_1T_2T_3 T_4 T_5
13	adolescent	1872	0.05	5	T_1 T_2 T_3 T_4 T_5
14	chronic	1863	0.05	5	T_1 T_2 T_3 T_4 T_5
15	dose	1819	0.05	5	T_1 T_2 T_3 T_4 T_5
16	abuse	1691	0.04	5	T_1 T_2 T_3 T_4 T_5
17	maternal	1543	0.04	5	T_1 T_2 T_3 T_4 T_5
18	psychiatric	1462	0.04	5	T_1 T_2 T_3 T_4 T_5
19	acid	1451	0.04	5	T_1 T_2 T_3 T_4 T_5
20	protein	1396	0.04	5	T_1 T_2 T_3 T_4 T_5
22	hormone	1209	0.03	5	T_1 T_2 T_3 T_4 T_5
23	gene	1193	0.03	5	T_1 T_2 T_3 T_4 T_5
24	serum	1156	0.03	5	T_1 T_2 T_3T_4 T_5
25	congenital	1120	0.03	5	T_1 T_2 T_3 T_4T_5
26	insulin	1115	0.03	5	T_1 T_2 T_3 T_4 T_5
27	bowel	1076	0.03	5	T_1 T_2 T_3 T_4 T_5
28	diabetes	1041	0.03	5	T_1 T_2 T_3 T_4 T_5
29	tissue	1037	0.03	5	T_1 T_2 T_3 T_4 T_5
30	muscle	1020	0.03	5	T_1 T_2 T_3 T_4 T_5
31	renal	1016	0.03	5	T_1 T_2 T_3 T_4 T_5
32	dysfunction	993	0.03	5	T_1 T_2 T_3 T_4 T_5
33	oral	993	0.03	5	T_1 T_2 T_3 T_4 T_5
34	medication	991	0.03	5	T_1 T_2 T_3 T_4T_5
35	glucose	966	0.03	5	T_1 T_2 T_3 T_4 T_5

Number	Word	Frequency	%	Range	Subdisciplines
36	liver	946	0.03	5	T_1 T_2 T_3 T_4 T_5
37	defect	932	0.03	5	T_1 T_2 T_3 T_4 T_5
38	gastric	910	0.03	5	T_1 T_2 T_3 T_4 T_5
39	receptor	885	0.025	5	T_1 T_2 T_3 T_4 T_5
40	plasma	882	0.025	5	T_1 T_2 T_3 T_4 T_5
41	artery	881	0.025	5	T_1 T_2 T_3 T_4 T_5
42	trauma	854	0.020	5	T_1 T_2 T_3 T_4 T_5
43	platelet	838	0.020	5	T_1 T_2 T_3 T_4 T_5
44	thyroid	793	0.020	5	T_1 T_2 T_3 T_4 T_5
45	alcohol	779	0.020	5	T_1 T_2 T_3 T_4 T_5
46	cardiac	767	0.020	5	T_1 T_2 T_3 T_4 T_5
47	impair	759	0.020	5	T_1 T_2 T_3 T_4 T_5
48	fetus	742	0.020	5	T_1 T_2 T_3 T_4 T_5
49	surgery	742	0.020	5	T_1 T_2 T_3 T_4 T_5
50	hypertension	739	0.020	5	T_1 T_2 T_3 T_4 T_5

Note. The words in the brackets are the headwords which did not meet the criteria to include in the MAWLOT. Instead, the family members which met the criteria to include in the MAWLOT were inserted in the list.

Subdisciplines: T_1= Internal Medicine, T_2= Psychiatry, T_3= Pediatrics, T_4= Surgery, T_5= Obstetrics and Gynecology

The 50 final words of the MAWLOT with their frequency and range are presented by

Table 5.2 follow.

Table 5.2: The fifty final words of the medical academic word list of textbooks with their frequency and range

Number	Word	Frequency	%	Range	Subdisciplines
1	neuron	100	0.003	4	T_1 T_2 T_3 T_4
2	oxidation	100	0.003	5	T_1 T_2 T_3 T_4 T_5
3	tubular	100	0.003	5	T_1 T_2 T_3 T_4 T_5
4	erythropoietin	101	0.003	3	T_1 T_3 T_5
5	ovulate	101	0.003	4	T_1 T_2 T_3 T_5
6	scar	101	0.003	4	T_1 T_2 T_3 T_5
7	(fungus) fungal	102	0.003	4	T_1 T_2 T_3 T_4
8	smear	102	0.003	4	T_1 T_2 T_3 T_5
9	(face) facial	103	0.003	5	T_1 T_2 T_3 T_4 T_5
10	glomerular	103	0.003	5	T_1 T_2 T_3 T_4 T_5
11	occult	103	0.003	5	T_1 T_2 T_3 T_4 T_5
12	remission	103	0.003	3	T_1 T_2 T_3
13	glycogen	104	0.003	3	T_1 T_2 T_3
14	inject	104	0.003	4	T_1 T_2 T_3 T_5
15	inpatient	104	0.003	4	T_1 T_2 T_3 T_5
16	graft	105	0.003	5	T_1 T_2 T_3 T_4 T_5
17	hematopoietic	105	0.003	3	T_1 T_3 T_5
18	nucleus	105	0.003	5	T_1 T_2 T_3 T_4 T_5

Number	Word	Frequency	%	Range	Subdisciplines
19	supplement	105	0.003	4	T_1 T_2 T_3 T_5
20	toxin	105	0.003	4	T_1 T_2 T_3 T_5
21	uterus	105	0.003	4	T_1 T_2 T_3 T_5
22	dilate	106	0.003	4	T_1 T_2 T_3 T_5
23	involuntary	106	0.003	5	T_1 T_2 T_3 T_4 T_5
24	surveillance	107	0.003	4	T_1 T_2 T_3 T_5
25	excretion (excrete)	108	0.003	5	T_1 T_2 T_3 T_4 T_5
26	febrile	108	0.003	5	T_1 T_2 T_3 T_4 T_5
27	infraction (infract)	108	0.003	5	T_1 T_2 T_3 T_4 T_5
28	manic	108	0.003	4	T_1 T_2 T_3 T_5
29	peritonitis	108	0.003	4	T_1 T_3 T_4 T_5
30	soluble	108	0.003	5	T_1 T_2 T_3 T_4 T_5
31	ventricle	108	0.003	3	T_1 T_3 T_5
32	inborn	109	0.003	4	T_1 T_2 T_3 T_5
33	pica	109	0.003	3	T_1 T_2 T_3
34	alveolar	110	0.003	5	T_1 T_2 T_3 T_4 T_5
35	cortex	110	0.003	5	T_1 T_2 T_3 T_4 T_5
36	folic	110	0.003	5	T_1 T_2 T_3 T_4 T_5
37	local	110	0.003	4	T_1 T_2 T_3 T_5
38	oxidase	110	0.003	5	T_1 T_2 T_3 T_4 T_5
39	physiology	110	0.003	5	T_1 T_2 T_3 T_4 T_5
40	carbohydrate	112	0.003	5	T_1 T_2 T_3 T_4 T_5
41	incise	112	0.003	3	T_1 T_3 T_5
42	pelvis	112	0.003	5	T_1 T_2 T_3 T_4 T_5
43	sensation	113	0.003	5	T_1 T_2 T_3 T_4 T_5
44	contraception	114	0.003	4	T_1 T_2 T_3 T_5
45	mania	115	0.003	3	T_1 T_2 T_3
46	mesenteric	115	0.003	4	T_1 T_2 T_3 T_4
47	pathological	115	0.003	3	T_1 T_2 T_5
48	clot	116	0.003	4	T_1 T_2 T_3 T_4 T_5
49	extracellular	116	0.003	5	T_1 T_2T_3 T_4 T_5
50	superior	116	0.003	5	T_1 T_2 T_3 T_4 T_5

Note. The words in the brackets are the headwords which did not meet the criteria to include in the MAWLOT. Instead, the family members which met the criteria to include in the MAWLOT were inserted in the list.

Subdisciplines: T_1= Internal Medicine, T_2= Psychiatry, T_3= Pediatrics, T_4= Surgery, T_5= Obstetrics and Gynecology

5.2.2 Correlation

The word families in the Medical Academic Word List of Textbooks (MAWLOT) occurred in a wide range of subdisciplines in the Medical Academic Corpus. Of the 459

word families in the list, 269 (58.7%) covered all 5 subdisciplines, 148 (32.3%) covered 4 subdisciplines, and 42 (9%) covered 3 subdisciplines. In total, 417 word families (91%) in the MAWLOT occurred in 4 or more of the subdisciplines under study. To understand the relationship between the frequency and range of the word families included in the MAWLOT the test of the *linear correlation coefficient (r)* was done. In statistics this test is used to assess what relationship, if any, exists between two variables. It measures the direction and strength of the relationship and this trend is represented by a correlation coefficient, most often represented symbolically by the letter '*r'* when it is measured in a sample, while the symbol for Pearson's correlation is 'ρ' when it is measured in the population. According to Altman (1991, cited in Mukaka, 2012), use and misuse of correlation is so common that some statisticians have wished that the method had never been devised. In the present study, I used *'r'* to represent Pearson's correlation because I deal exclusively with samples. There are other types of variable measurement tools such as Kendall's Rank or Spearman's Rank, but these measure different types of relationships and cannot be used as an alternative to the linear measurement tool.

The linear correlation coefficient is a mathematical representation of the mathematical relationship between two values or sets of data. It is the statistical measurement of degree to which the change in one of the measurements affects the change in another set of measurements. The *linear correlation coefficient (r)* is sometimes referred to as the Pearson product-moment correlation coefficient with the value of between $-1 < r < +1$ to measure the strength and the direction of a linear relationship between two variables. This means that the greater the absolute value of a correlation coefficient, the stronger the linear relationship. In other words, a positive correlation shows that if one variable gets bigger, the other variable tends to get bigger as well. On the other hand, a negative correlation means that if one variable gets bigger,

the other variable tends to get smaller. Therefore, an *r* value of exactly -1 indicates a perfect negative fit. But, the weakest linear relationship is indicated by a correlation coefficient equal to 0. If there is no linear correlation or a weak linear correlation, *r* is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables. It is also important to know that the Pearson product-moment correlation coefficient only measures linear relationships. Therefore, a correlation of 0 does not mean zero relationship between two variables; rather, it means zero *linear* relationship.

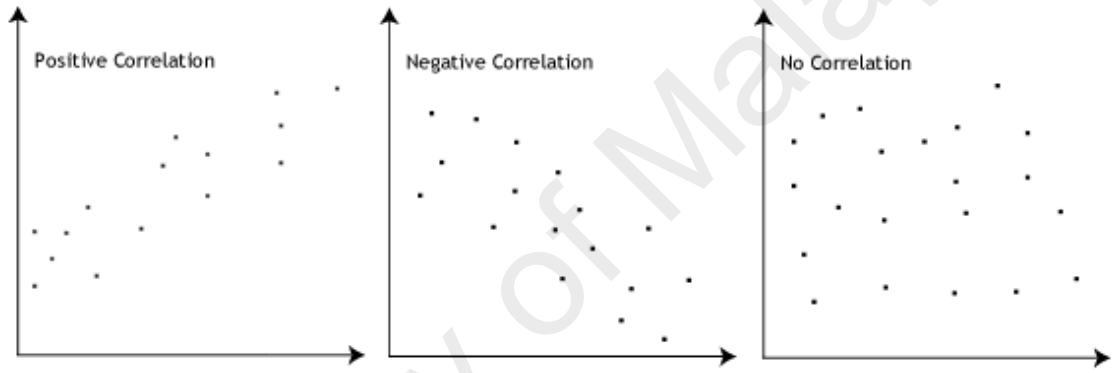


Figure 5.1: The different ranges of values from +1 to -1 of the Pearson correlation coefficient (Laerd Statistics, 2017)

There are several formulas that can be used to compute Pearson's correlation. Some formulas make more conceptual sense whereas others are easier to actually compute. For example, the Pearson's correlation between the variables X and Y is computed by dividing the sum of the xy column (Σxy) by the square root of the product of the sum of the x^2 column (Σx^2) and the sum of the y^2 column (Σy^2). The resulting formula is:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

There are also many software packages, which can be used to compute a correlation coefficient. For example, the correlation coefficient is part of the statistical formulas available in the Microsoft Excel. The advantage of using the Formulas tab to initiate the function is that the Function Arguments window opens up when a function is selected, and the window helps to enter the required parameters for the function.

I used the Microsoft Excel (2013), as the latest version of the program software in the time of starting the study, to measure the correlation coefficient. The Excel is the most popular spreadsheet available today. The part of the reason for its popularity is the fact that Excel comes standard with hundreds of functions and formulas. These functions and formulas allow users to quickly and effectively manipulate and analyze data within the spreadsheet, without having to necessarily understand how the complex calculations that form the basis of these formulas work.

The present study shows a positive correlation between the frequency and range of the word families included in the MAWLOT (r_s= *0.05, p= 0.000*), which means that when the frequency of the MAWLOT's words is increased and gets bigger, the range of the words tends to be increased and gets bigger as well (see Appendix H for the complete test of the *linear correlation coefficient*). The study of Wang et al. (2008) to establish a medical academic word list of research articles also reports a positive relationship between the frequency and range of words included in the word list (r_s= *0.753, p = 0.000*).

5.2.3 The medical technical words of the Medical Academic Corpus

A list of medical technical words was also retrieved to increase the students' knowledge in reading medical texts. These words are discipline-specific words (domain-specific academic words or content-specific words) (Hiebert & Lubliner, 2008) and are particularly unique in the area of medicine with just one meaning. According to

Coxhead (2000), applying a list of technical words in a particular field with an appropriate word list helps to increase the comprehensibility of academic texts in the field. These medical technical words, which met the criteria for inclusion in the word list ($F \geq 100$, $R \geq 3$), were added to the study as a separate list. Table 5. 3 shows the medical technical words of the study.

Table 5.3: The medical technical words of the medical academic corpus

Number	Word	Frequency	%	Range	Subdisciplines
1	acidemia	179	0.005	4	T_1 T_3 T_4 T_5
2	acidosis	202	0.006	5	T_1 T_2 T_3 T_4 T_5
3	adenocarcinoma	120	0.003	4	T_1 T_3 T_4 T_5
4	anastomosis	192	0.005	3	T_1 T_4 T_5
5	anatomic	194	0.005	4	T_1 T_3 T_4 T_5
6	anemia	385	0.011	5	T_1 T_2 T_3 T_4 T_5
7	anesthesia	175	0.005	5	T_1 T_2 T_3 T_4 T_5
8	aureus	212	0.005	3	T_3 T_4 T_5
9	biliary	222	0.006	4	T_1 T_2 T_3 T_4
10	bilirubin	157	0.005	4	T_1 T_3 T_4 T_5
11	biologic	169	0.005	3	T_1 T_3 T_4
12	carcinoid	100	0.003	4	T_1 T_2 T_3 T_4
13	cesarean	529	0.015	4	T_1 T_2 T_3 T_5
14	cns	187	0.005	5	T_1 T_2 T_3 T_4 T_5
15	colorectal	196	0.005	3	T_1 T_3 T_4
16	corticosteroids	187	0.005	5	T_1 T_2 T_3 T_4 T5
17	crohn	358	0.01	4	T_1 T_2 T_3 T_4
18	cushing	128	0.003	3	T_1 T_2 T_4
19	cytokines	156	0.005	5	T_1 T_2 T_3 T_4 T_5
20	dehydrogenase	107	0.003	4	T_1 T_3 T_4 T_5
21	distention	114	0.003	4	T_1 T_2 T_3 T_4
22	doppler	103	0.003	4	T_2 T_3 T_4 T_5
23	ectopic	178	0.005	5	T_1 T_2 T_3 T_4 T_5
24	erectile	106	0.003	4	T_1 T_2 T_3 T_4
25	erythematosus	187	0.005	5	T_1 T_2 T_3 T_4 T5
26	esophageal	512	0.014	5	T_1 T_2 T_3 T_4 _5
27	esophagus	494	0.014	5	T_1 T_2 T_3 T_4 T_5
28	estrogen	134	0.003	5	T_1 T_2 T_3 T_4 T_5
29	etiology	478	0.014	5	T_1 T_2 T_3 T_4 T_5
30	fecal	105	0.003	5	T_1 T_2 T_3 T_4 T_5

31	femoral	131	0.003	4	T_1 T_3 T_4 T_5
32	FSH	111	0.003	3	T_1 T_2 T_3
33	gastroesophageal	123	0.003	4	T_1 T_2 T_3 T_4
34	gonadotropin	124	0.003	5	T_1 T_2 T_3 T_4 T_5
35	hemoglobin	138	0.003	4	T_1 T_3 T_4 T_5
36	HIV	263	0.007	5	T_1 T_2 T_3 T_4 T_5
37	hydrops	103	0.003	3	T_1 T_3 T_5
38	hyperglycemia	131	0.003	5	T_1 T_2 T_3 T_4 T_5
39	hypoglycemia	306	0.009	5	T_1 T_2 T_3 T_4 T_5
40	hypothalamic	223	0.006	4	T_1 T_2 T_3 T_4
41	idiopathic	119	0.003	5	T_1 T_2 T_3 T_4 T_5
42	inguinal	251	0.007	4	T_1 T_3 T_4 T_5
43	insipidus	162	0.005	4	T_1 T_3 T_4 T_5
44	ischemia	101	0.003	5	T_1 T_2 T_3 T_4 T_5
45	MRI	143	0.004	5	T_1 T_2 T_3 T_4 T_5
46	neutropenia	155	0.004	3	T_1 T_3 T_4
47	neutrophils	118	0.003	3	T_1 T_3 T_4
48	nonspecific	136	0.004	5	T_1 T_2 T_3 T_4 T_5
49	osmolality	108	0.003	4	T_1 T_3 T_4 T_5
50	physiologic	166	0.004	5	T_1 T_2 T_3 T_4 T_5
51	psychosocial	114	0.003	3	T_1 T_2 T_3
52	Tumors (tumor)	341	0.010	4	T_1 T_2 T_3 T_4
53	vasopressin	203	0.006	4	T_1 T_2 T_3 T_4

Note. Subdisciplines: T_1= Internal Medicine, T_2= Psychiatry, T_3= Pediatrics, T_4= Surgery, T_5= Obstetrics and Gynecology

5.2.3.1 The effect of medical technical words to increase the MAWLOT's coverage of medical textbooks

Applying a list of technical words in a particular field with an appropriate word list helps to increase the vocabulary knowledge of learners, according to Lessard-Clouston (2013). According to the findings yielded by the descriptive phase of the study, in addition to the Medical Academic Word List of Textbooks (MAWLOT), a list of medical technical words (53 words) that met the criteria for inclusion in the word list was also retrieved. As said above, these discipline-specific words, domain-specific academic words or content-specific words (Hiebert & Lubliner, 2008) are particularly unique in the area of medicine, as most of them have only one meaning. To test the hypothesis of Coxhead (2000), who posited that applying a list of technical words in a

particular field besides an appropriate word list helps to increase the vocabulary knowledge of academic texts in the field, the medical technical words were examined in a random sample of texts (approximately 750-850 words or two pages long) from each subcorpus (see Appendix F for excerpts from texts in each discipline of the Medical Academic Corpus). Next, we added the medical technical words to the MAWLOT (we called it MAWLOT+), and re-examined the coverage of the MAWLOT in the texts (see Appendix G). Employing the medical technical words together with the words included in the MAWLOT increased the texts' coverage of the MAWLOT in all subcorpora of the vocabulary of the corpus (see Table 5.4).

Table 5.4: Coverage of the subcorpora samples by the medical academic word list of textbooks and the medical academic word list of textbooks besides the technical words (%)

Subcorpus	Coverage of the Medical Academic Word List of Textbooks	Coverage of the Medical Academic Textbooks Besides the Medical Technical Words
Internal medicine	12.06	16.27
Obstetrics &Gynecology	10.39	10.91
Psychiatry	9.23	12.23
Pediatrics	9.48	9.84
Surgery	14.37	15.14

5.2.4 Frequent collocations of the words included in the MAWLOT

Having a concordance list of the words included in the word list will help to study the words in their context. Therefore, to observe the behaviour of the most frequent words included in the medical academic word list of textbooks (MAWLOT) in context, this study has presented the collocation/phraseological patterns of the words in Chapter 6.

5.2.5 Frequent collocations of the medical technical words of the Medical Academic Corpus

Similar to the words included in the MAWLOT, the most frequent collocations of the words included in the Medical Technical Word List were retrieved from the Medical Academic Corpus to study words in multiple contexts or observe words' behaviour with a greater number of word encounters. Chapter 6 explains extensively the frequent collocations of the medical technical words of the study.

5.2.6 Differences in occurrence of words across subdisciplines of medicine

The third question was whether the lexical items selected for the MAWLOT occur with different frequencies in the subcorpora of medicine, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry and surgery. Processing the data with Text Lex Compare (TLC) v.3, from the Research part of the Compleat Lexical Tutor (CLT) reveals the differences in occurrences of words included in the MAWLOT across the medical subcorpora. According to the findings, the MAWLOT appears to be slightly advantageous for pediatrics texts (12.51%) and similar to the coverage of internal medicine (12.14%). Then, there are the coverage of surgery (11.78%), obstetrics and gynecology (10.74%), and psychiatry (9.21%), as Table 5.5 shows.

Table 5.5: Coverage of the subcorpora of the medical academic corpus by the medical academic word list of textbooks

Subcorpus	Coverage (%)
Internal medicine	12.14
Obstetrics & Gynecology	10.74
Psychiatry	9.21
Pediatrics	12.51
Surgery	11.78

5.3 Findings of the Evaluation Phase (Research Questions 4 - 6)

The findings of the three last research questions (Research Questions 4 - 6) answers the evaluation phase of the present study, which seeks to understand the total coverage of the Medical Academic Corpus beyond the GSL (West, 1953) and the AWL (Coxhead, 2000), the coverage of another medical academic corpus, and the comparison of the MAWLOT with AWL and MAWL (Wang et al., 2008) in terms of types, tokens, running words of corpora, and the total coverage of each word list retrieved from the corpora.

5.3.1 Coverage of the Medical Academic Corpus beyond the GSL (West, 1953) and the AWL (Wang et al., 2008)

The first step in evaluating any word list is measuring the coverage of included words in the corpus (Coxhead, 2000). According to Coxhead (2000), the coverage of the AWL's word families is 10% of the vocabulary in the Academic Corpus. Coxhead reports that the coverage of the disciplines of arts, law and science by the AWL in her study was similar, i.e. 9.3%, 9.4%, and 9.1%. The discipline of commerce had the highest coverage with 12.0%. In comparison, the coverage of the MAWL (Wang et al., 2008) has been reported 12.24% of the vocabulary included in the corpus of medical research articles by the researchers. In the present study, the findings show that the vocabulary frequency distribution of the MAWLOT among the subdisciplines of the corpus is different in each of the five subcorpora (i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery). The cumulative text coverage of the MAWLOT's word families across subdisciplines of medicine in the Medical Academic Corpus is 11.27%, which is higher than the coverage of the AWL across the Academic Corpus (10%).

As to the lexical text coverage of each subdiscipline, the coverage of the MAWLOT for the subdiscipline of internal medicine is 12.14%, similar to the coverage of the subdiscipline of pediatrics, i.e. 12.51%. The coverage of the MAWLOT's word families across subdisciplines of the obstetrics and gynecology, psychiatry, and surgery is 10.74%, 9.21%, and 11.78% (see Table 5.5). It is also important to note that the MAWLOT is shorter in length with 459 word families in comparison of the MAWL with 623 words. It is thus clear that the MAWLOT should serve medical students much better than the AWL and the MAWL (explained in section 7.6). The MAWLOT is offering substantially more coverage while containing significantly fewer items, which makes the process of learning easier. Thus, given the students' low vocabulary size (Hajiyeva, 2015), Hylandand and Tse (2007) believe that reducing the learning effort requires a very important point to consider in the development of academic word lists, which has been done in developing the MAWLOT. This study also provides evidence for the need to include, in discipline-specific vocabulary lists such as the MAWLOT, the medical technical words with a significantly higher frequency and special meanings in the discipline can improve the vocabulary knowledge of the intended learners by increasing the coverage of the word list.

5.3.2 Coverage of another medical academic corpus

Obviously, a frequency-based word list of a particular corpus expects to cover the corpus well. Therefore, to have a real test of any word list it has to be examined in similar texts of a different collection (Coxhead, 2000). As said before (see section 4.8.5), to consider whether the MAWLOT maintains the high coverage over medical academic texts of the corpus, I compiled a second balanced corpus by using the same criteria and sources to select texts. The Second Corpus was made of the same subcorpora, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery.

The findings showed that the MAWLOT's word families occur differently, but with a good coverage in the subcorpora of the Second Corpus (10.48%), which is more than the coverage of the Academic Corpus by the AWL, i.e. 8.5% (Coxhead, 2000). The coverage of the MAWLOT for the subcorpora of internal medicine is the highest coverage at 14.45%. The coverage for the subcorpora of obstetrics and gynecology, psychiatry, and surgery is very similar -- 9.88%, 9.98%, and 9.53%, respectively. The lowest coverage of the MAWLOT by the Second Corpus belonged to the subcorpora of surgery at 8.58%. Table 5.6 shows these findings in details.

Table 5.6: Coverage of the second corpus by the medical academic word list of textbooks

Subcorpus	**Coverage (%)**
Internal medicine	14.45
Obstetrics & Gynecology	9.88
Psychiatry	9.98
Pediatrics	8.58
Surgery	9.53

5.3.3 Comparison of the MAWLOT with the AWL (Coxhead, 2000) and the MAWL (Wang et al., 2008)

The Academic Word List (AWL) (Coxhead, 2000) was developed from the Academic Corpus with 3,513,330 running words (tokens) of written academic text by examining the frequency and range of words outside the first 2,000 most frequent words in English, i.e. the General Service List (GSL) (West, 1953). The corpus to develop the AWL included 3,110 different types of words by including the disciplines of arts, commerce, law, and sciences. The resulting list (AWL) contained 570 word families with coverage of about 10% of the vocabulary of the Academic Corpus, as Coxhead mentions. AWL highlights the most frequent words for study by university students.

Also, to establish the medical academic word list (MAWL) of research articles, Wang et al. (2008) included 1,093,011 running words in a corpus consisting of different subdisciplines of medicine. The types of the words included in the corpus have not been reported. The resulting word list (MAWL) contained 623 word families with coverage of 12.24%. According to Wang et al., of the 623 word families in the MAWL, 104 (16.69%) words occurred in all 32 subject areas, while 321 (51.52%) words occurred in 25 or more subject areas. In addition, the frequency and the range of the word families included in the MAWL were positively correlated (r_s*= 0.753, p = 0.000*).

Regarding the present study, the Medical Academic Word List of Textbooks (MAWLOT) was compiled from a balanced corpus with 3,513,378 running words (very similar to the running words included in the Academic Corpus of Coxhead's (2000) study) with 2,659 different types of the words. The Medical Academic Corpus contains the major subdisciplines of medicine, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery. The final word list of the study, i.e. Medical Academic Word List of Textbooks (MAWLOT) consists of 459 word families, with a good coverage of 11.27% to represent the most frequent words of medical textbooks, which occurred in a wide range of the Medical Academic Corpus. The findings show that, of the 459 word families in the list, 58.7% covered all 5 subdisciplines, 32.3% covered 4 subdisciplines, and 9% covered 3 subdisciplines. In total, 91% in the MAWLOT occurred in 4 or more of the subdisciplines under study (r_s*= 0.05, p= 0.000*). Comparison MAWLOT with the AWL and the MAWL is illustrated in Table 5.7

Table 5.7: Comparison of medical academic word list of textbooks with the academic word list and the medical academic word list

Characteristic	Medical Academic Word List of Textbooks	Academic Word List (Coxhead, 2000)	Medical Academic Word List (Wang et al. 2008)
Running words (tokens)	3,513,378	3,513,330	1,093,011
Types	2,659	3,110	--------
Word families (total)	459	570	623
Coverage of Academic Corpus (%)	11.27	10	12.24

5.4 Summary

In this chapter, the findings of the study were presented in two main parts: the descriptive phase and the evaluation phase. Regarding the descriptive phase, the findings of Research Question 1 revealed the Medical Academic Word List of Textbooks (MAWLOT) with 459 words. In addition, a list of the medical technical words, which contained 53 word families was identified. The list of the most frequently collocation/phraselogical patterns of the MAWLOT's words was also retrieved (see chapter 6). Next, the differences in occurrence of word families included in the MAWLOT across the subdisciplines of Medical Academic Corpus was tested. According to the findings, the MAWLOT appears to be slightly advantageous for pediatrics texts (12.51%), similar to the coverage of internal medicine (12.14%).

Furthermore, the findings of the evaluation phase showed that the coverage of the MAWLOT of the vocabulary of the Medical Academic Corpus is 11.27%, which is a good coverage and also more than the coverage of the AWL (Coxhead, 2000) (10%) by the Academic Corpus. Next, to have a real test of MAWLOT, I examined the word list in similar texts of a different collection. The findings showed that the MAWLOT's word families occur differently, and with a good coverage in the subcorpora of the Second Corpus, i.e. 10.48%, which is also more than the coverage of the second Academic Corpus by the AWL (8.5%) (Coxhead, 2000).

In comparing the MAWLOT with the AWL (Coxhead, 2000) and the MAWL (Wang et al. 2008) in terms of types, tokens and running words of the corpora, besides the total coverage of the word lists of the vocabulary in the corpora, it was found that the developed word list of the present study (MAWLOT) is more useful for medical students in EAP/ EMP teaching and learning, which is discussed in Chapter 7.

CHAPTER 6: COLLOCATION/PHRASEOLOGICAL PATTERNS OF THE MEDICAL ACADEMIC WORD LIST OF TEXTBOOKS

6.1 Introduction

This chapter considers the patterns that are associated with particular lexical items and that are observable from investigation of the Medical Academic Corpus of the study. To have a view of the chapter, the role of collocations in teaching language is considered, first. Next, the most frequent collocation/phraseological patterns of the words included in the Medical Academic Word List of Textbooks (MAWLOT) is presented. The chapter ends with the discussions related to the findings.

6.2 The utility of teaching collocation in L2

Conducting different research and presenting different definitions of 'collocation' have shown the importance of collocation in teaching English (see section 3.10). For example, Lehecka (2015) states that "collocation in particular has become a fundamental concept in usage-based studies in many linguistic fields, most notably lexical syntax and semantics" (p. 2). Shabani (2016) also points to the importance of teaching collocation and mentions that "If L2 words are not properly taught, students soon will face communicational problems" (p. 21). Woolard (2000) believes that collocation is an important category of lexical patterning and it is becoming a recognized unit of description in language teaching courses and materials. Wray (2002) and Nesselhauf (2003) state that collocation helps to enhance both accuracy and fluency in L2. In addition, in his study Liu (2000) observed that teaching collocations repetitively causes learners to use them more correctly and appropriately to express what they mean. Nation (2000) states that collocational knowledge is in "the essence of language knowledge" (p. 318), and the process of learning words requires knowledge of those collocates.

In addition, it should be noticed that collocations are examples of lexical units (Shabani, 2016), which shape the components of the lexical approach. For example, Richards and Rodgers (2001) state that "the building blocks of language learning and communication are not grammar, functions, notions, or some other unit of planning and teaching but lexis, that is, word and word combinations" (p. 132). Martyńska (2004) also states that "the meaning of a word is determined by the co-occurring words. Consequently, lexis is considered to be independent and separable from grammar. Thus, a part of the meaning of a word is the fact that it collocates with another word" (p. 1). It should also be noted that in collocations "words are learned and used in context" (Sadeghi, 2009, p. 104).

Furthermore, there are some researchers who emphasize on teaching and learning collocation as they know it is absolutely crucial to language learners (e.g. Ghezelseflou & Seyedrezaei, 2015). These researchers believe that most EFL learners only learn the definitions of English words, so their passive vocabulary cannot be easily reconstructed into acceptable chunks or natural and meaningful sentences (Sarvari, 2016). Likewise, Wu (1996) states that collocations "cannot be spontaneously acquired" (p. 469). Therefore, lexical collocation could play a role as "one way to make passive vocabulary active" (p. 468).

In addition, collocations play a significant role in communicative competence (Sadeghi, 2009), as the competency in using collocations leads to a decrease in language learners' anxiety (Kuo, 2009) to solve lexical ambiguities (Leffa, 1998). Also, according to Martyńska (2004), knowing how words combine into chunks (collocation) is necessary because if words are not learned in chunks, learners will not be able to achieve the native-like level of proficiency. Wu (1996) goes further and states that "if learners are not familiar with collocations, they can only generate awkward sentences with traces of native-language interference" (p. 9).

Furthermore, the role of frequency has been considered as an important criterion for selecting what to focus on in collocations studies (Sarvari, 2016). Sadeghi (2009) states that "the immediate implication is a need for beginning learners consciously to learn high-frequency collocations and for intermediate and advanced EFL learners to learn less frequent ones (in addition to highly frequent ones)" (p. 116). Other factors include "range of use, difficulty, teachability, and suitability for the age and background of the learners" (Shin & Nation, 2008, p. 346). According to Walker (2011), by studying the collocations associated with a group of so-called synonyms it is often possible to identify slight but significant differences in the meaning of the words in the group. It noted that collocations are stored in the mental lexicon and accessed holistically (Sarvari, 2016). The mechanism which determines the constituency is the frequency with which items occur together in natural, connected speech, that is, the collocational frequency (Sosa & MacFarlane, 2002). In addition, Hsu (2007) found a strong relationship between the frequency of lexical collocations and online writing fluency. Related to the present study is Gledhill's (2000) investigation of the relation between collocations of high frequency words in medical abstracts and articles and with prototypical phraseology of the genre.

Regarding teaching collocations, concordancing has been recognized as a desirable approach by some researchers (e.g. Zahar, Cobb & Spada (2001)). These researchers state that teaching concordancing raises a student-centered instruction by supplying L2 learners with multiple exposures to new items and collocations. They also point out to learners' frequency of exposure to the collocations to be learned. One of the important roles of concordance lines is to illustrate the patterning of language (Hunston & Francis, 2000) by including the criterion of frequency to select the pattern. The important role of teaching patterns in language teaching is discussed in section 7.3.1.2.

6.3 Frequent collocation/phraseological patterns of the words of the MAWLOT

According to Hunston and Francis (2000), "A pattern can be identified if a combination of word occurs relatively frequently" (p. 37). As said in the Methodology chapter, in the present study we follow the procedure of Hunston and Francis (2000) to find the most frequent patterns of the MAWLOT's words. In the first step, 20 statistically significant top words (0.25% - 0.04 %) were selected from the list of the 50 top words included in the MAWLOT to be considered for finding their frequent collocation/phraseological patterns. Next, 50 concordance lines were selected at random from all retrieved lines which contained the word one was examining to find the frequent and particular patterns revealed by each subcorpus.

The following part presents the significant patterns of the MAWLOT's words resulted from considering the concordance lines of each subcorpus. Consequently, the discussion on the resulted patterns is presented in section 6.5. The sample lines have been reduced to 1-2 lines to represent the patterns of the words (see section 4.3.1). The selected 20 top words are: *patients*, *disorder*, *diagnosis*, *therapy*, *syndrome*, *fetal*, *cell*, *clinical*, *infant*, *drug*, *cancer*, *adolescent*, *chronic*, *dose*, *abuse*, *maternal*, *psychiactric*, *acid*, and *protein* as follow.

1. patients

a. internal medicine

The selected concordance lines of the subcorpus of internal medicine reveal the patterns **PL-N with n** and **in PL-N with n**. The pattern **PL-N with n** means that the plural noun *patients* is followed by the particular preposition *with* and a noun. The pattern **in PL-N with n** means that the plural noun is preceded by the specific preposition *in*:

PL-N with n

1 lar perturbations associated with OSAHS. In obese patients with a BMI greater than 35 kg/m2,

4 iazem or verapamil in hemodynamically compensated patients with a contraindication to?-blockers; an

in PL-N with n

3 evidence is seen, it is diagnostic. In patients with a compatible clinical syndrome, excl

25 best used as a confirmatory test in patients with a history suggestive of neurocardiog

b. obstetrics and gynecology

PL-N with n

4 testing, both antibodies were found. In general, patients with LAC have higher levels of

in PL-N with n

15 A: Antiepileptic drug use and birth rate in patients with epilepsy\x97a population-based

26 ents (Tsokos, 2001). Some autoantibodies produced in patients with lupus are shown in Table 54-1.

c. pediatrics

in PL-N with n

1 is indicated using 1 to 2 J/kg. In patients with a complaint of palpitations, it is

36 Mycobacterium avium-complex (MAC) in patients with AIDS Clarithromycin, azithromycin, r

d. psychiatry

PL-N with n

38 relationships occur, some therapists do not accept patients with AIDS into a group unless they

49 ive disorders. The compromised metabolism of many patients with anorexia can put them at high

in PL-N with n

9 to prevent seasonal major depressive episodes in patients with a history of SAD. Smoking Cessation

26 frontal lobe or right hemisphere hypometabolism in patients with ADHD, compared with normal controls.

e. surgery

PL-N with n

23 of HNPCC. In as many as 50% of patients with a family history that clearly demons

47 o assist in the pretreatment planning for patients with a lung mass. Invasive staging of

in PL-N with n

22 and for about 15% of such cancers in patients with a family history of colorectal cancer

36 ing and nonsecreting neoplasms (Fig. 39-18). In patients with a history of malignancy, metastatic

2. disorder

a. internal medicine

The noun *disorder* is followed by a range of prepositions such as: ***as, at, but, by, during*** to present the pattern **N prep.** Also, it is followed by **and** in the pattern **N and,** and preceded by an adjective in the pattern **adj N**:

N prep

28	for any disorder depends on understanding the	disorder at	the genetic level. The characteristics
32	e the clinical appearance of a myeloproliferative	disorder by	several years, so patients older than

N and

15	, if a woman's father has the	disorder and	her mother is a carrier, she
19	physician needs to be familiar with the	disorder and	major potential complications. For ex

adj N

8	than would be predicted for the simple	disorder,	an additional process has to be present
14	ngs (Table 32-1). Dependence is the more severe	disorder	and frequently is associated with physiol

b. obstetrics and gynecology

N with n

6	1/127 95% Neurodegenerative lysosomal storage	disorder with growth	failure, marked psychomotor r
8	s that may manifest: late-onset neurodegenerative	disorder with tremor	and ataxia, particularly in m

c. pediatrics

In the subcorpus of pediatrics, the selected concordance lines at random reveals the pattern **N is characterized by**, which means that the noun is followed by the specific verb *is*, the specific past participle *characterized,* and the specific preposition *by*. The other patterns revealed by the subcorpus are: **adj N**, **N and**, **N is pp**, **N is n**, and **N is adj**:

N and

38	ental disorder, schizophrenia, or other psychotic	disorder and	are not better accounted for by
40	sleep disorder, such as sleep-onset association	disorder and	difficulty falling asleep, do not nee

adj N

2	motor movements indicate a tic. Body dysmorphic	disorder,	a delusional fixation on appearance, can
15	schizophrenia, or attention-deficit/hyperactivity	disorder	[ADHD]). Having a first-degree relative w

N is characterized by

11	, and pheochromocytoma. Substance-induced anxiety	disorder is characterized by	impairing symptoms of
15	> autistic disorder autistic disorder autistic	disorder is characterized by	marked impairment in

N is pp

9 abnormal reproduction, and some have both. This disorder is caused by mutations in the KAL

17 subsequent development of panic disorder. Panic disorder is described in all racial groups. Lifeti

N is n

1 nality disorder. Obsessive-compulsive personality disorder is a character structure involving preocc

30 An example of this type of mitochondrial disorder is MELAS (mitochondrial encephalomyopathy

N is adj

8 occur on different chromosomes (Trans), and the disorder is benign because few individuals of Afri

50 ciated with psychosis. The etiology of autistic disorder is unknown. There is an increased risk

d. psychiatry

N is associated with

41 attening of schizophrenia itself. This depressive disorder is associated with an increased risk of

43 d and adolescent conduct disorders. When conduct disorder is associated with ADHD, a trial of

N is n

9 Disorder The necessary feature of bipolar I disorder is a history of a manic or

12 in Table 27-15. Differential Diagnosis Dysthymic disorder is a mood disorder characterized by great

N is adj

24 allows clinicians to specify whether the pain disorder is acute or chronic, depending on whether

adj N

10 of trust may be manifested by dysthymic disorder, a depressive disorder, or a sense of

12 cause: primary, or caused by another mental disorder, a general medical condition, or substance

e. surgery

N is characterized by

2 hrinks dramatically in Huntington's disease. This disorder is characterized by rigidity, on which is

N is pp

4 present with the same intensity, a neuromuscular disorder is suspected. When a patient is found

adj N

7 also when the onset of a mental disorder is rapid. Otherwise, little is known abou

11 the source(s), and whether the current disorder is the first episode for the patient.

N prep

6 characterize this disease. ITP is principally a disorder of increased platelet destruction mediate

N and

4 is. Other possibilities are infection, autoimmune disorder and reaction from chronic, contained rup

23 made during the workup for a motility disorder and the diverticulum is found incidental

3. diagnosis

e. internal medicine

In the subcorpus of internal medicine, the selected concordance lines at random reveal the patterns **N of n**, **N is made**, **N of acute n** and **N of acute MI**. The pattern N **of acute MI** means that the noun is followed by the specific preposition *of*, the specific adjective *acute,* and the specific noun *MI*:

N of n

7 n in the general dialysis population. Diagnosis of ACKD is relatively straightforward an

33 irregularity of the RR wave makes the diagnosis of AF relatively easy. Aberrant conducti

N is made

1 of symptoms and the time when the diagnosis is made. At the time of diagnosis, 90%

2 ite requires early surgical consultation once the diagnosis is made. The classic J wave on

N of acute n

9 of cough is useful for considering the diagnosis of acute bronchitis. Patients usually se

26 (see later). Diagnosis As part of the diagnosis of acute respiratory failure, the physic

N of acute MI

14 mes Clinical Manifestations Traditionally, the diagnosis of acute MI has rested on the

20 ity substantially improves the specificity of the diagnosis of acute MI with only a modest

b. obstetrics and gynecology

N of adj

1 , Weiner CP, Williamson RA, et al: Prenatal diagnosis of fetal hyperthyroidism using funipunct

21 g continues and hypovolemia becomes significant. Diagnosis of Ectopic Pregnancy Laboratory Tests

N is based on

2 ness, and coma (Chapa and colleagues, 2003). The diagnosis is based on clinical symptoms and the

N is confirmed by

1 vels decrease as placental epitopes are secreted. Diagnosis is confirmed by the failure of an

2 elop multiple sclerosis within 15 years. Clinical diagnosis is confirmed by MR imaging and cerebrosp

N is made

47 that termination generally is indicated when the diagnosis is made. Certainly, before 24 weeks, con

23 40 percent of women are asymptomatic, and the diagnosis is made during routine sonographic exami

c. pediatrics

The patterns **N of n**, **N of adj**, **N of acute n**, **N is based on**, **N is confirmed by**, **N is established**, **N is made by**, and **N is adj** are the particular patterns revealed by the subcorpus of pediatrics:

N of adj

45	has been discontinued. Table 80-2. Differential	Diagnosis of Atopic Dermatitis Congenital disord
48	or from a storage disease. The differential	diagnosis of autoimmune neutropenia includes syste

N of n

27	Behaviour Checklist, are not specific for the	diagnosis of ADHD. The physical examination is e
39	Table 105-1. Definition of Acute Otitis Media A	diagnosis of AOM requires History of acute ons

N of acute n

19	to cats. The most reliable test for	diagnosis of acute EBV infection is the IgM

N is based on

5	ndrome, and symptoms result from this deficiency.	Diagnosis is based on the amino acid pattern
8	by precipitation of alpha chain tetramers. The	diagnosis is based on an elevation of hemoglobin

N is confirmed by

19	Kawasaki syndrome, serum sickness, and drug rash.	Diagnosis is confirmed by serologic testing for Ig
27	emphysema is present in the neck. The	diagnosis is confirmed by radiograph. Treatment is

N is established

34	i-inflammatory therapy with salicylates after the	diagnosis is established, and bed rest. Additional
39	Antibiotics often are begun before a definitive	diagnosis is established, which complicates the ab

N is made by

6	heard in the left second interspace. Definitive	diagnosis is made by cardiac catheterization, but
7	s highly overrated. Laboratory Diagnosis. The	diagnosis is made by identifying the precise amount

N is adj

31	has its own specific treatment, and proper	diagnosis is crucial. At least 25% of children wit
42	. If no cause can be determined, the	diagnosis is idiopathic precocious puberty; this c

d. psychiatry

N of adj

36	45. Solomon L, Reeves WC. Factors influencing the	diagnosis of chronic fatigue syndrome. Arch Intern

N of n

12	impairs social or occupational functioning. The	diagnosis of alcohol or drug use in adolescents

19 patient agreed on the latter option. A diagnosis of anxiety disorder was made, which allo

N is made

4 attention span may be present before the diagnosis is made. The virus can be present

7 seen by a psychiatrist before the correct diagnosis is made. Word production, as well as

N is established

1 a persistent vegetative state and once the diagnosis is established DNR and do not intubate

e. surgery

The subcorpus of surgery reveals the patterns **N of acute**, **N of n**, **N of adj**, **N of adrenal insufficiency**, **N of acute abdominal pain**, **N of acute appendicitis**. The patterns **N is based on**, **N is confirmed by**, **N is made**, **N is made by**, and **N is established** are the other particular patterns revealed by the subcorpus:

N of n

15 arrest have the highest mortality. Because the diagnosis of AAA was unknown in more than 70%

20 vated. An undetectable ACTH level establishes the diagnosis of ACTH-independent Cushing's syndrome a

N of adj

5 ribed, are malignant. Fine-needle aspiration for diagnosis of a new SPN in a patient

16 of spontaneous reflux lends support to the diagnosis of abnormal gastroesophageal reflux. The

N of acute

39 uired. Transbronchial biopsy may be performed for diagnosis of acute rejection. Acute rejection usual

N of adrenal insufficiency

42 As is true for most endocrine disorders, diagnosis of adrenal insufficiency depends on main

45 luation of critically ill patients. [15] After the diagnosis of adrenal insufficiency has been made,

N of acute abdominal pain

1 first published his classic paper on the diagnosis of acute abdominal pain in 1921. [6] An e

2 Sevier differential diagnosis the differential diagnosis of acute abdominal pain is extensive. Co

N of acute appendicitis

8 and signs and symptoms consistent with a diagnosis of acute appendicitis. At exploration, t

10 90% and a specificity of 80% to 90% for the diagnosis of acute appendicitis among patients with

N is based on

3 an elevated C-reactive protein level. The diagnosis is based on the patients' history and

9 more diffuse when caused by Giardia lamblia. Diagnosis is based on biopsy specimens or scraping

N is confirmed by

16 g, tenesmus, and occasionally abdominal pain. The diagnosis is confirmed by a careful history and

17 1 infusion is required to stabilize the patients. Diagnosis is confirmed by echocardiography; other

N is made

23 ut childhood and early adulthood. Frequently, the diagnosis is made incidentally. In other patients,

30 isease represents a surgical emergency. After the diagnosis is made, operation is performed in an

N is made by

35 the differential diagnosis for hyperkalemia. The diagnosis is made by demonstrating elevated serum

39 artburn, chest pain, and regurgitation are noted. Diagnosis is made by manometry. IEM is defined

N is established

2 is it usually the only presenting symptom. Diagnosis is established by small bowel contrast s

4 right ventricle and pulmonary artery. After the diagnosis is established, patients should undergo

4. Therapy

a. internal medicine

The subcorpus of internal medicine reveals the patterns **N for acute**, **N for adj**, **N for n**, and **N is pp**:

N for adj

19 for chelation therapy. The role of chelation therapy for asymptomatic individuals with mercury

28 , Liong ML, Yuen KH, et al: Terazosin therapy for chronic prostatitis/chronic pelvic pai

N for n

14 lsely negative. Treatment The cornerstones of therapy for AKI are rapid recognition and correcti

25 utility of this approach in developing targeted therapy for breast and ovarian cancer. Accuracy

N for acute

6 mbocytopenic purpura Immune globulin, intravenous Therapy for acute episodes Botulism Trivalent A, B

11 rombolytic therapy (see Treatment). Treatment Therapy for acute PE overlaps substantially with t

N is pp

2 These evaluations should ensure that when medical therapy is added, changed, or deleted, these adjus

23 ndotracheal and tracheostomy tubes. If aggressive therapy is considered appropriate, ligation of the

b. obstetrics and gynecology

N for adj

29 (1999), the addition of heparin to antimicrobial therapy for septic pelvic thrombophlebitis did not

30 gues 2002; Schummer and Schummer 2002). Principal therapy for toxic shock is supportive, while allow

N for n

12	ds. Twin-Twin Transfusion Syndrome (TTTS) Laser	therapy for TTTS has been the most commonly
26	d mastitis. Treatment provided that appropriate	therapy for mastitis is started before suppuration

N is adj

2	IA trachomatis. Thus, azithromycin or doxycycline	therapy is appropriate empirical therapy. Andrew
23	r increasing -HCG, additional surgical or medical	therapy is necessary. Medical Management with Met

N is pp

12	warts during pregnancy is not always necessary.	Therapy is directed toward minimizing treatment to
21	en with frequent migraine headaches, prophylactic	therapy is indicated (American College of Obstetri

c. pediatrics

The subcorpus of pediatrics reveals the patterns **N for n**, **N for adj**, **N is pp**, and **N is adj**:

N for adj

2	ly exchange transfusion) usually are indicated in	therapy for acute chest syndrome. Incentive spirom
13	, onset 3-4 mo of age may require folate	therapy for chronic hemolysis; hematocrit usually

N for n

4	, which are infused into the patient. Gene	therapy for ADA deficiency has not been successful
7	the possible exception of M3 AML). Induction	therapy for AML usually consists of cytarabine, da

N is pp

1	of respiratory fatigue. In other patients, oxygen	therapy is administered using appropriate methods.
10	plantation as soon as induction and consolidation	therapy is completed. Gemtuzumab ozogamicin, a dru

N is adj

8	restricted in the intake of leucine, glycine	therapy is beneficial through enhancement of the f
16	reases urinary potassium losses. Disease-specific	therapy is effective in many of the genetic

d. psychiatry

In the subcorpus of psychiatry, the noun *therapy* is followed by the specific verb *is*, and a noun to reveal the patter **N is n**. The other patterns revealed by the subcorpus are **N is pp**, **N is adj**, **N for adj**, and **N for n**:

N for adj

8	Cognitive behavioural and attachment based family	therapy for anxious adolescents: Phase I and II
39	roup psychotherapy has been the recommended group	therapy for pubertal children who do not have

N for n

2	) is typically the second line of pharmacologic	therapy for ADHD in children who do not

4 tic-induced akathisia probably effective Adjunctive therapy for alcohol withdrawal and other substance

N is n

2 s. Couples (Marital) Therapy Couples or marital therapy is a form of psychotherapy designed to

N is pp

42 , a child who is to receive intravenous therapy is helped by acting out the procedure

46 use. If the findings are positive, anticonvulsant therapy is initiated or appropriate surgery is pro

N is adj

5 as part of the therapeutic process. Combined therapy is a particular treatment modality, not a

11 hniques, including massage. Chelation Chelation therapy is a traditional medical procedure used to

e. surgery

In the subcorpus of surgery the most frequently patterns revealed by the subcorpus are **N for n**, **N for adj**, **N is pp**, **N is adj**, and **N is n**:

N for adj

25 gnificantly during the past 5 decades of surgical therapy for congenital cardiac disease. Initial su

44 tive comparison between percutaneous and surgical therapy for hepatic abscess, but case series sugge

N for n

9 ated complications. Angioplasty Versus Medical Therapy for Angina N=1018 F/U =2.7 years

15 ASD closure before school age. The standard therapy for ASDs for the past 50 years has

N is pp

2 -up of sclerotherapy for hepatic cysts. Surgical therapy is achieved by fenestration or unroofing o

24 uccessful healing is roughly 72% if no additional therapy is employed. If H2-receptor antagonists ar

N is adj

13 or determining the utility of hormone replacement therapy is complex, however, because estrogen-repl

26 less than 30 mL/m2, prognosis after surgical therapy is excellent. Progressive systolic dysfunc

N is n

6 up to 90% of patients. The preferred endoscopic therapy is argon plasma coagulation (Fig. 46-7).

5. syndrome

a. internal medicine

The noun *syndrome* reveals the patterns, in which the noun is followed by a specific verb *is*, and a noun as **N is n**. The other pattern revealed by the subcorpus is **N is pp prep**:

N is n

1	rlying cause of the disturbance. Hyperventilation	syndrome is a diagnosis of exclusion, but patients
9	1?-hydroxysteroid dehydrogenase. The cause of the	syndrome is activation of the mineralocorticoid re

N is pp prep

17	131-5) that it is understandable why the uremic	syndrome is attributed to toxicity. Symptoms inclu
24	tract involved in the Wolff-Parkinson-White	syndrome is composed of fibers containing INa and

b. obstetrics and gynecology

The subcorpus of obstetrics and gynecology reveals the particular patterns of **N is n**, **N is pp**, **N is pp prep**, **N is associated with**, **N is caused by**, and **N is characterized by**:

N is n

2	ulation proteins (Hall and associates, 1980). The	syndrome is a phenocopy of chondrodysplasia puncta

N is pp

18	tation, and respiratory distress. Generally, the	syndrome is considered to be mild and self-
19	y Syndrome (APS) By international consensus, the	syndrome is diagnosed on the basis of clinical

N is pp prep

10	the maternal 15q11-q13. In 2 percent, the	syndrome is caused by paternal uniparental disomy,
31	follow surgical procedures or immunizations. The	syndrome is thought to be immune-mediated from

N is associated with

5	Taylor and associates, 2009). It appears that the	syndrome is associated with decreased endothelial
6	Plasma Volume Expansion because the preeclampsia	syndrome is associated with hemoconcentration dire

N is caused by

8	tical deletions. Another possibility is that each	syndrome is caused by a different contiguous gene
9	r. In approximately 70 percent of cases, Angelman	syndrome is caused by microdeletion or disruption

N is characterized by

13	of events that leads to the preeclampsia	syndrome is characterized by a host of abnormaliti
16	he antiphospholipid antibody syndrome (APS). This	syndrome is characterized by recurrent arterial or

c. pediatrics

The patterns **N is n**, **N is adj**, **N is pp prep**, **N is associated with**, **N is caused by**, and **N is characterized by** are revealed by the subcorpus of pediatrics:

N is n

2	letion, 20% have UPD of chromosome 15. Angelman	syndrome is a condition with a characteristic faci

5 nt dysfunction (see Table 150-8). The acute chest syndrome is a vaso-occlusive crisis within the

N is pp prep

34 ic population. The diagnosis of myofascial pain syndrome is based on the presence of multiple

N is adj

9 nomegaly and anemia are present. Crigler-Najjar syndrome is a serious, rare, permanent deficiency

N is associated with

15 leukocyte adhesion deficiency, whereas hyper-IgE syndrome is associated with cold abscesses, eczema

18 somal dominant inheritance pattern. Laurence-Moon syndrome is associated with spastic paraplegia, ho

N is caused by

23 oximately 25% of cases of Marfan syndrome. Marfan syndrome is caused by many mutations. Virtually ev

25 high mortality rates. Hantavirus cardiopulmonary syndrome is caused by Sin Nombre virus, which

N is characterized by

26 ately 1 in 10,000 live-born infants, Prader-Willi syndrome is characterized by hypotonia of prenatal

33 failure, or hepatorenal syndrome, may occur. This syndrome is characterized by low urine output, azo

d. psychiatry

N is n

1 Abnormal Swallowing Syndrome Abnormal swallowing syndrome is a condition during sleep in which

31 of their symptoms. Some clinicians believe this syndrome is neurasthenia, a diagnosis used primary

N is pp

13 by the World Health Organization (WHO). The syndrome is believed to occur in about 30 percent

39 otency agents are used. If neuroleptic malignant syndrome is suspected, the DRA should be stopped

N is pp prep

36 perceive it as a whole). Balint's syndrome is seen in bilateral parieto-occipital le

N is adj

2 to isolate from the brain. Gulf War syndrome is a controversial condition with inflamm

e. surgery

The patterns most frequently revealed by the subcorpus of surgery are **N is pp**, **N is adj**, and **N is caused by**:

N is pp

19 . The treatment of complications of Peutz-Jeghers syndrome is directed mainly at the complication of

47 toms or develops pathologic fractures. Albright's syndrome is suspected if these lesions are multipl

N is adj

13 eurologic disorders. The underlying cause of this syndrome is bacterial overgrowth in stagnant areas

39 onsumptive coagulopathy known as Kasabach-Merritt syndrome is rare but well described. [33] LFTs and

N is caused by

1 syndrome, also termed ACTH-independent Cushing's syndrome is caused by autonomous adrenal cortisol

6. fetal.

a. internal medicine

The pattern **prep cells of ADJ origin** is revealed by the subcorpus of internal medicine:

prep cells of ADJ origin

1 the fetus, examining DNA in cells of fetal origin, analyzing chromosomes in fetal cells

4 clinical benefit after therapy with cells of fetal origin that has now lasted up to 10

b. obstetrics and gynecology

The subcorpus of obstetrics and gynecology reveals the patterns **ADJ heart n**, **ADJ heart rate**, **ADJ heart rate abnormalities**, and **ADJ heart rate acceleration**. The pattern **ADJ heart n** reveals that ADJ is followed by the specific noun *heart*, and a noun. In the pattern **ADJ heart rate**, the adjective is followed by the specific noun *heart*, and another specific noun *rate*:

ADJ heart n

9 actions. Descriptive characteristics of baseline fetal heart activity include rate, beat-to-beat

16 Maternal corticosteroid administration to treat fetal heart block is controversial. Shinohara and

ADJ heart rate

36 The technique requires internal monitoring of the fetal heart rate and special equipment to process

5 1F quiet sleep. During state 2F, the fetal heart rate baseline bandwidth increased appr

ADJ heart rate abnormalities

33 in advance because these fetuses commonly have fetal heart rate abnormalities during labor. Indee

36 hese children with matched controls. Intrapartum fetal heart rate abnormalities did not distinguish

ADJ heart rate acceleration

44 (1975) introduced the nonstress test to describe fetal heart rate acceleration in response to fetal

49 infarction in 93 percent. Thus, the lack of fetal heart rate acceleration, when not due to

c. pediatrics

The subcorpus of pediatrics reveals the patterns **ADJ alcohol syndrome**, **ADJ heart rate abnormalities**, **ADJ heart rate**, and **during ADJ life**:

ADJ heart rate

3	the neonatal CNS. If recovery of the	fetal heart rate	occurs as a result of
4	onium-stained fluid maternal fever/amnionitis	fetal heart rate	patterns (distress) Scalp pH

ADJ alcohol syndrome

15	forms of brain developmental disorder, including	fetal alcohol syndrome	and Down syndrome, and in
23	characteristic palmar crease pattern is seen in	fetal alcohol syndrome	. Genitalia Genitalia sh

ADJ heart rate abnormalities

2	sly, as should uterine contractions during labor.	Fetal heart rate abnormalities	may indicate basely

during ADJ life

1	from the pulmonary artery to the aorta	during fetal life	. Failure of the normal closure
4	precursors to migrate into the distal bowel	during fetal life	. The aganglionic distal segment

d. psychiatry

The subcorpus of psychiatry reveals the pattern **ADJ alcohol syndrome**:

ADJ alcohol syndrome

2	seizures. The incidence of infants born with	fetal alcohol syndrome	is about 0.5 per 1,000 live
6	congenital rubella), prenatal exposures (e.g.	fetal alcohol syndrome	), and specific chromosomal

e. surgery

ADJ heart

1	The sinus venosus, which empties into the	fetal heart	, ultimately becomes the hepatocardiac

during ADJ development

3	musculature to unite in the midline during	fetal development.	The umbilical vessels may be sp
8	the hematopoietic function is usually lost during	fetal development	, the spleen continues to function

during ADJ life

1	ormalities; this analysis may be performed during	fetal life	through amniocentesis. The chromosomal
4	is from pulmonary artery to aorta during	fetal life	). After the first breath at delivery

6. cell

a. internal medicine

The subcorpus of internal medicine reveals the patterns **T N activation**, **B N activation**, **Sickle N anemia**, **Sickle N n**, **germ N tumors**, **squamous N carcinomas**, and **squamous N carcinomas of the n** for the noun *cell*:

T N activation

7	the secondary wave that ensues after T-	cell activation, can also induce expression of HLA
13	IL-2 with a resulting decrease in T-	cell activation. Cyclosporine also inhibits the pr

B N activation

8	Low-affinity Fc receptor for IgE, B-	cell activation CD25 Activated T and B cells,
16	hematopoiesis IL-4 T cells, mast cells B-	cell activation, IgE switch, inhibition of TH1 cell

Sickle N anemia

3	disease can occur with oral contraceptive use,	sickle cell anemia and other hemoglobinopathies, p
9	neuropathies in diabetes, analgesic nephropathy,	sickle cell anemia, systemic lupus erythematosus (

Sickle N n

14	ac origin Decreased oxygen supply Anemia,	sickle cell disease Hypoxemia (e.g. sleep apn
22	idney and bone marrow transplant nephropathy, and	sickle cell nephropathy. Males who present with hy

germ N tumors

3	include thymomas, lymphomas, teratomas and other	germ cell tumors, substernal thyroid goiters, lipo

squamous N carcinomas

1	, include sunburn, skin cancers (basal cell and	squamous cell carcinomas and, to a lesser extent,
2	Bowen's disease, basal cell ca rcinoma, and	squamous cell carcinomas (Chapter 214). Lung can

squamous N carcinomas of the n

3	of p53 mutations that are associated with	squamous cell carcinomas of the lung, head, and

b. obstetrics and gynecology

Sickle N anemia

2	of Southeast Asian or African ancestry; and	sickle-cell anemia for people of African, Mediterr
4	ochromatosis Homocystinuria Phenylketonuria	Sickle cell anemia Tay-Sachs disease Thalassem

Sickle N n

8	ic sclerosis, primary pulmonary hypertension, and	sickle cell disease. Obstet Gynecol 111

c. pediatrics

The subcorpus of pediatrics reveals the pattern **n N n**, which means that the noun is preceded by a noun and followed by another noun. In the other patterns, **N** is combined

and preceded by the specific nouns *sickle*, *T*, and *germ* to make the patterns **sickle N disease**, **sickle N anemia**, **T N activation**, and **germ N tumors**:

sickle N anemia

2 ticulocytopenia, suggests bone marrow failure Sickle cell anemia and -thalassemia appear as fet

20 ical Manifestations and Treatment. A child with sickle cell anemia is vulnerable to life-threateni

n N n

14 rare and usually results from an islet cell adenoma. Children with this condition charact

26 testes may be seen. Patients with T cell ALL are frequently older boys (8 to 10 years)

sickle N disease

14 iabetes mellitus Malnutrition Uremia Sickle cell disease Zinc deficiency Multiple carboxyl

18 ical antibiotic treatment. Children with sickle cell disease have impaired splenic function and pr

T N activation

2 anscription of early cytokine genes necessary for T cell activation. These agents may be used

germ N tumors

5 ectodermal tumors (including medulloblastoma) and germ cell tumors are sensitive to chemotherapy; gl

9 relatively common in children and include mature germ cell tumors/hamartomas, hemangiomas or other

d. psychiatry

The particular pattern revealed by the subcorpus of pediatrics is **n N n**, which means that the noun is preceded by a noun and followed by another noun:

n N n

2 LFT, liver function test; WBC, white blood cell, ADHD, attention-deficit/hyperactivity disord

6 of the "stored nutrient" in the nerve cell (neuron). This depletion resulted from stress

e. surgery

The subcorpus of surgery reveals the patterns **islet N tumors** and **germ N tumors**, which means that the noun is preceded by the specific nouns *islet* and *germ*, and followed by the plural noun *tumors*. The other patterns revealed by the subcorpus are **n N n**, **squamous N carcinoma**, **squamous N carcinoma of n**, **adj N carcinoma**, **Sickle N anemia**, and **Sickle N n**:

n N n

3 cystic carcinoma (colloid carcinoma), signet ring cell carcinoma, adenosquamous carcinoma, 127naplastic

islet N tumors

25 will await multi-institutional trials. All islet cell tumors except insulinomas (10%) and GRFomas

32 in an autosomal dominant fashion. Because islet cell tumors in patients with MEN 1 are always

germ N tumors

7 diastinum, particularly thymomas, lymphomas, germ cell tumors, and primary carcinomas, can have rema

14 related to recurrent pneumonitis. Although germ cell tumors are rare, the diagnosis can be

squamous N carcinoma of n

42 cultures have a high rate of squamous cell carcinoma of the esophagus, not related to

44 colitis. Extraintestinal cancer, such as squamous cell carcinoma of the vulva and anal canal

squamous N carcinoma

17 helial Neoplasia the condition of anal squamous cell carcinoma in situ (Bowen's disease) was

24 to an 8% chance of developing carcinoma. Squamous cell carcinoma is the most common type identified

adj N carcinoma

7 nosquamous carcinoma, anaplastic carcinoma, giant cell carcinoma, and sarcomatoid carcinoma are cons

8 related to sun exposure and include basal cell carcinoma and SCC. Keratoacanthoma is a benig

Sickle N anemia

1 are life-threatening disorders in children with sickle cell anemia and thalassemia. In these condi

4 subject to splenic sequestration and destruction. Sickle cell anemia is the result of the

Sickle N n

1. hemoglobin variants, such as hemoglobin C or sickle cell –thalassemia. Sickle cell disease res

6. clinical

a. internal medicine

The subcorpus of internal medicine reveals the particular patterns **adj ADJ**, **ADJ and adj**, **ADJ manifestatons prep**, and **ADJ manifestations of**:

adj ADJ

14 ropriate. The work-up should include appropriate clinical laboratory evaluation to identify der

ADJ and adj

11 on, patients typically develop the characteristic clinical and hemodynamic features of classic distr

16 ent perturbations in neuronal function. Thus, the clinical and neurologic consequences of hypertonic

ADJ manifestations prep

8 rom cerebral, pulmonary, and other vascular beds. Clinical manifestations after the traumatic even

11 crease after prolonged storage of blood products. Clinical manifestations although most cases of T

ADJ manifestations of

1. mimics some, but not all of the clinical manifestations of absolute hypovolemia. T

9 cal manifestations of the underlying disease, the clinical manifestations of hypervolemia depend on

b. obstetrics and gynecology

adj ADJ

17 ent consists primarily of monitoring the maternal clinical and laboratory conditions as well as fetal

ADJ and adj

11 Third-degree obstetric perineal tear: Long-term clinical and functional results after primary repa

ADJ manifestations prep

7 syndrome toxin-1\x97TSST-1\x97causes the clinical manifestations by provoking profound endo

ADJ manifestations of

19 (1995), are that there is no correlation between clinical manifestations of disease and complement

20 as with most organ-specific autoimmune disorders, clinical manifestations of endocrinopathies result

c. pediatrics

adj ADJ

34 is rarely severe or associated with significant clinical bleeding. Congenital (TORCH [toxoplasmosi

ADJ and adj

7 tconceptual age and accompanied by characteristic clinical and radiographic findings that correspond

11 sis, keratomalacia, and a permanent corneal scar. Clinical and subclinical vitamin A deficiencies ar

ADJ manifestations

10 ross the duodenum, causing partial obstruction. Clinical manifestations about 60% of children wi

36 apnea also may lead to core pulmonale. Clinical manifestations and diagnostic Studies T

d. psychiatry

adj ADJ

1 and is based on the established structural, clinical, and postulated neurochemical characteris

e. surgery

ADJ manifestations of

24 primary HPT were typically seen with the clinical manifestations of hypercalcemia, includin

9. acute

a. internal medicine

The patterns **ADJ respiratory acidosis**, **ADJ respiratory alkalosis**, **ADJ respiratory distress syndrome**, **ADJ renal failure**, and **ADJ adj** are revealed by the subcorpus of internal medicine:

ADJ adj

5 lready administered. Intoxication may result from acute, chronic, or acute-on-chronic exposure. A

22 lungs is more likely to produce an acute, severe pneumonitis. Once material has been

ADJ respiratory acidosis

1 derlying cause and ensuring adequate ventilation. Acute respiratory acidosis can be very dangerous,

3 ding headaches and papilledema. Other findings in acute respiratory acidosis include signs of catech

ADJ respiratory alkalosis

6 patient with hypophosphatemia as a result of acute respiratory alkalosis do not require treatm

10 by a decrease in bicarbonate levels. In acute respiratory alkalosis, one of the primary me

ADJ respiratory distress syndrome

11 Diffuse alveolar filling process similar to the acute respiratory distress syndrome diffuse alveol

24 opacities consistent with acute lung injury and acute respiratory distress syndrome. The CT scan i

ADJ renal failure

7 kidney dysfunction, but they only rarely cause acute renal failure. Acute renal failure may be

12 ocalcemia, metastatic calcification, hypotension, acute renal failure, and arrhythmias, as well as

b. obstetrics and gynecology

ADJ adj

2 normal brain histopathology was not observed with acute, lethal asphyxia. Moreover, 43 percent of br

4 matomyositis and Polymyositis these are uncommon acute, subacute, or chronic inflammatory diseases

c. pediatrics

ADJ adj

35 rfere with normal growth and development, whereas acute, severe changes in pH can be fatal.

41 to an associated pyarthrosis. (From Gutman LT: Acute, subacute, and chronic osteomyelitis and pyo

ADJ abdominal pain

3 rapid. Only a few children presenting with acute abdominal pain actually have a surgical emer

10 ion. Table 126-1 lists a diagnostic approach to acute abdominal pain in children. Important clues

an ADJ abdomen

2 . Abdominal pain occurs frequently and can mimic an acute abdomen. The presence of polyuria, despite

d. psychiatry

ADJ adj

1 precipitated by a variety of drugsRule out acute abdominal disease, acute psychiatric episode

28 lithium however, lamotrigine does not have any acute antimanic effects. Chemistry Lamotrigine i

ADJ stress disorder

4 ive disorder (OCD), generalized anxiety disorder, acute stress disorder, and posttraumatic stress di

7 , PTSD, and acute stress disorder. PTSD and acute stress disorder have the nature of the

ADJ stress disorder PTSD

16 group of trauma spectrum disorders that includes acute stress disorder, PTSD, and somatization diso

20 the diagnosis of adjustment disorder, PTSD, and acute stress disorder. PTSD and acute stress disor

e. surgery

an ADJ abdomen

6 of nearly all diseases presenting as an acute abdomen are presented. A must read for

10 be bloody. Abdominal pain may mimic an acute abdomen. Bacteremia should be treated by adm

prep ADJ abdomen

2 in a large series of patients with acute abdomen. A good review of this important

7 [19] [20] [21] Many of the most common causes of acute abdomen are readily identified by CT scanning

prep ADJ abdominal pain

2 ematoma is an uncommon condition characterized by acute abdominal pain and the appearance of an

4 and treatment of the gravid female with acute abdominal pain and is now routinely employed

10. infant

a. internal medicine

The patterns **N mortality** and **adj N** are the most frequently patterns revealed by the subcorpus of internal medicine:

N mortality

1 renal tract malformations are important causes of infant mortality and of morbidity in older children

2 ons, stillbirths, low birthweights, and increased infant mortality. At mean blood lead levels above

adj N

9 normal result does not guarantee a normal infant. If the result of the test is

14 risk of having a low-birth-weight infant to that of a woman who has

b. obstetrics and gynecology

N mortality

1 defects are currently the leading cause of infant mortality and account for 20 percent of deal

5 meconium, only a few are linked to infant mortality. In an investigation from Parklan

adj N

12 of neonatal intensive care for a preterm infant. Home Healthcare Many clinicians believe

17 or delivery of a live or stillborn infant. In some locales, the obstetrical history i

N mortality rate

7 racial/ethnic minorities has persisted, and the infant mortality rate in African-American women is

8 reported a remarkably low combined perinatal and infant mortality rate of 0.63 percent. Fetal Inju

c. pediatrics

The subcorpus of pediatrics reveals the patterns **N mortality**, **N is pp**, **the N is –ing**, **the N is adj**, and **adj N**:

N mortality

4 ning, and appropriate nutrition and health care. Infant mortality increases as the mother's level

6 . New Hampshire and Minnesota had the lowest infant mortality rates in the U.S., whereas

adj N

1 newborns. When the patient is a young infant, a careful dietary history is crucial. A

15 of cyanosis is required for every cyanotic infant after prompt administration of oxygen, with

N is pp

9 to prevent perinatal HIV transmission; if an infant is confirmed as HIV-infected while receivin

22 Signs After the general appearance of the infant is evaluated, the examination should procee

the N is -ing

12 desired, with care taken to ensure the infant is consuming a variety of other foods

15 mouth. Although the growth rate of the infant is decreasing, energy needs for activity in

the N is adj

41 surgical intervention as needed later when the infant is stable, and a laparotomy can be

44 than the mother's titer, if the infant is symptomatic, or if the mother has

d. psychiatry

adj N

5 constipation that food is not an unborn infant. Erikson often turned to play, which, alon

28 . In fact, the movements of an anencephalic infant, who completely lacks a cerebral cortex, ma

N is pp

4 modify the instinctual drives with which an infant is born. The id, however, should not

12 the inner from the outer world; mother– infant is perceived as a single fused entity.

e. surgery

N is prep

1 mesenteric rotation and fixation, and hence the infant is at risk for mesenteric volvulus with

11. drug

a. internal medicine

The patterns **N is pp prep** and **N interactions** are revealed by the subcorpus of internal medicine:

N is pp prep

6 from a quantification of how fast the drug is cleared from the circulation. Drug clear

17 transplantation and autoimmune disease. Often this drug is used in combination with other agents

N interactions

4 compounded by the heightened potential for drug- drug interactions. If more than three drugs are

7 increased toxicity. Because of multiple potential drug interactions that can both raise and lower

b. obstetrics and gynecology

N is pp prep

9 used currently to identify teratogenicity after a drug is released for clinical use include follow-

N interactions

1 as the efficacy of hormonal contraception. These drug interactions are detailed by the U.S.

4 pills but seek attention to exclude pregnancy. Drug Interactions Combination oral contraceptives

c. pediatrics

N is pp prep

5 given at least twice daily. If the drug is stopped for more than 48 hours, the

N interactions

2 a low dietary intake, malabsorption, or vitamin- drug interactions. Deficiency can develop within a

3 and achieve therapeutic levels in these sites. Drug interactions must be considered when multiple

d. psychiatry

The particular pattern revealed by the subcorpus of psychiatry is **N interactions**:

N interactions

1 in DRA pharmacokinetic interactions. Other common drug interactions affect the absorption of the DRA

5 to serum proteins and has no significant drug interactions. As with donepezil, galantamine

e. surgery

IV N abuse

1 sickle cell disease), urinary tract infection, IV drug abuse, and AIDS. About 70% of splenic abscess

4 s occurs with endocarditis, osteomyelitis, and IV drug abuse. Splenic abscess may also occur as

12. cancer

a. internal medicine

The subcorpus of internal medicine reveals the following particular patterns:

prep n N in pl-n

17 all ages, whereas the estimates for breast cancer in women and for thyroid cancer in

N is pp prep

2 as well as from various forms of cancer is increased in heavily irradiated populati

N in n

3 , recent reports from Europe suggest excess brain cancer in cell phone users. These data are

5 r 18) synergistically increases the risk of lung cancer in cigarette smokers. Alcohol use interacts

b. obstetrics and gynecology

N is pp

21 ing and treatment of cervical carcinoma. Cervical cancer is staged clinically rather than surgically

N is pp prep

7 relative risk of cervical dysplasia and cervical cancer is increased in current COC users, but

N in n

5 : Breast-feeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. J

c. pediatrics

N is pp

9	lysis. Diagnostic Imaging When a diagnosis of	cancer is confirmed	, diagnostic imaging is importa

N is adj

2	valuating a child with suspected cancer. Childhood	cancer is rare	; only about 1% of new cancer

d. psychiatry

N is pp

1	therapies have gained popularity in recent years;	cancer is attributed	to a defective immune system,

N is adj

3	Child End-of-Life Care After accidents,	cancer is the second	most common cause of

e. surgery

The subcorpus of surgery reveals the patterns **N is adj n**, **N is pp prep**, **N in n**, and **N in patients with**:

N is pp prep

32	breast cancer. The excess risk for breast	cancer is concentrated in	women whose specimens sh
44	repeated with greater frequency. If an ampullary	cancer is discovered at	an early stage, pancreatod

N is adj n

8	, an imprint of elsevier lung cancer lung	cancer is a significant public	health problem in
19	. Local Excision Local excision of a rectal	cancer is an excellent operation	for a small

N in n

19	ressing metaplasia (SPEM) associated with gastric	cancer in Iceland	. Dig Dis Sci 2003; 48:431-441.
42	is shown in Figure 34-8, with invasive ductal	cancer in panel	A and invasive lobular cancer

N in patients with

47	with an 8-fold increased incidence of colorectal	cancer in patients with	polyps who do not
48	found will decrease the incidence of colon	cancer in patients with	HNPCC. However, there have

13. adolescent

a. internal medicine

The subcorpus of internal medicine reveals the pattern **ADJ n**:

ADJ n

5 Td if they have not received an adolescent Td booster. For patients who have recei

b. obstetrics and gynecology

ADJ pregnancy

10 and especially its intent in preventing unwanted adolescent pregnancy, a number of organizations, i

13 s represented a 9-percent decline since 2000, the adolescent pregnancy rate remains among the highes

c. pediatrics

ADJ n

33 a partnership way. Although the focus in adolescent care is on psychosocial issues, a gener

ADJ pregnancy

1 rths, and 29 abortions per 1000. The reduction in adolescent pregnancy since the peak year of 1990 i

3 be ruled out in all cases of adolescent pregnancy. When pregnancy is confirme

ADJ pl-n

21 victims of physical violence by their boyfriends; adolescent boys may become victimizers. Although s

d. psychiatry

ADJ n

4 tinence. Psychopharmacological interventions for adolescent alcohol and drug users are still in

45 adolescents with respect to the development of adolescent conduct disorder. Further studies are n

or ADJ antisocial behaviour

18 ctual functioning, academic problem, childhood or adolescent antisocial behaviour, and identity probl

20 malingering, adult antisocial behaviour, child or adolescent antisocial behaviour, borderline intellect

ADJ pl-n

38 , many programs offer exposure to child and adolescent cases. Traditionally, forensic child an

44 to be more aware of their sons' adolescent concerns than of their daughters' anxie

ADJ boys

31 on the street are informal and spontaneous. Adolescent boys frequently have contact with the l

32 ctor most strongly associated with violence among adolescent boys is growing up in a household

ADJ girls v

4 follicles, approximately once every 28 days. When adolescent girls reach SMR 3 to 4, ovarian follicle

6 with masturbation, and more than half of adolescent girls report masturbation. The balance

e. surgery

ADJ pl-n

4 dation may necessitate resection in pediatric and adolescent patients. Box 50-1 Ulcerative Colitis

14. chronic

a. internal medicine

The adjective *chronic* in the subcorpus of internal medicine reveals the particular patterns **prep ADJ obstructive lung disease**, **ADJ obstructive pulmonary disease**, **prep ADJ obstructive pulmonary disease**, **prep ADJ kidney disease**, **acute or ADJ**, and **prep acute or ADJ**:

prep ADJ obstructive lung disease

6 tis, synovium in rheumatoid arthritis, alveoli in chronic obstructive lung disease, and colonic epit

8 C, et al: Prevention of exacerbations of chronic obstructive lung disease with tiotropium,

ADJ obstructive pulmonary disease

19 pulmonary function Aggravation of asthma and chronic obstructive pulmonary disease Irritati

25 nditions (aortic atherosclerosis, renal function, chronic obstructive pulmonary disease, and coagula

prep ADJ obstructive pulmonary disease

15 can be misleading. Suppose a patient with chronic obstructive pulmonary disease and a histor

20 ions that suggest systemic diseases, ranging from chronic obstructive pulmonary disease, which may e

prep ADJ kidney disease

39 therapeutic options are much more limited for chronic kidney disease, it is less likely to

49 casts, the last-mentioned being seen in chronic kidney disease. RBC casts are rare in

acute or ADJ

2 in suggested texts or literature searches. In acute or chronic cases, information about what the

14 respiratory acidosis and respiratory alkalosis are acute or chronic. Such determination is based on

prep acute or ADJ

4 interstitial pneumonias, which are a subset of acute or chronic ILDs of unknown etiology, are

b. obstetrics and gynecology

acute or ADJ

2 97myeloid leukemias. They can be acute or chronic. Although adult leukemias are more prevale

c. pediatrics

The patterns **acute or ADJ**, **prep acute or ADJ**, **prep ADJ pancreatitis**, and **prep ADJ renal failure** are revealed by the subcorpus of pediatrics:

acute or ADJ

2 is infectious rhinitis, which may be acute or chronic. Acute infectious rhinitis, or the comm

4 slides Figure 78-1 Approach for managing acute or chronic asthma in infants and young children.

prep acute or ADJ

6 panel (electrolytes, liver enzymes, BUN) Signs of acute or chronic hepatic, renal, adrenal dysfuncti

prep ADJ pancreatitis

6 pain may be present. The causes of chronic pancreatitis include hereditary pancreatit

8 festations page 622 0 page623 Children with chronic pancreatitis initially present with recur

prep ADJ renal failure

17 Treatment The management of children with chronic renal failure and their complex problems r

29 much of the anemia formerly seen in chronic renal failure. Erythropoietin rapidly drai

d. psychiatry

The subcorpus of psychiatry reveals the patterns **prep ADJ fatigue syndrome** and **acute or ADJ**:

prep ADJ fatigue syndrome

3 viral infection frequently precedes the onset of chronic fatigue syndrome, no infectious agent has

11 on (CDC) defined specific diagnostic criteria for chronic fatigue syndrome. Since then, the disorder

acute or ADJ

1 to specify whether the pain disorder is acute or chronic, depending on whether the duration

e. surgery

The patterns of **prep ADJ pancreatitis**, **acute or ADJ**, **prep acute or ADJ**, and **prep ADJ renal failure** are revealed most frequently by the subcorpus of surgery:

prep ADJ pancreatitis

2 , rather than acute, pancreatitis. This type of chronic pancreatitis affects only the obstructed p

6 can develop. Many of the patients with chronic pancreatitis are alcoholics who, even before

acute or ADJ

4 etic blockade. Pseudo-obstruction may present in acute or chronic forms. The acute variety most HBeA

6 uring laparoscopic cholecystectomy. These include acute or chronic inflammation, obesity, and anatomic v

prep acute or ADJ

7 to treat and may be associated with acute or chronic obstructive symptoms. A type II

prep ADJ renal failure

2 hyperplasia. As many as 90% of patients with chronic renal failure have evidence of secondary H

5 strategies, there are pathophysiologic sequelae of chronic renal failure that serve as indications fo

15. dose

a. internal medicine

The subcorpus of internal medicine reveals the particular pattern **adj N of n**:

adj N of n

9 the morning dose follows the daily low dose of aspirin. A similar interaction has been

38 normal. The anticoagulant response to a standard dose of heparin varies widely among patients, such

b. obstetrics and gynecology

adj N of n

2 randomized trials have demonstrated that a single dose of an antimicrobial agent given at the

7 administration of a perioperative 2-g intravenous dose of cefazolin reduced rates of this morbidity,

c. pediatrics

adj N of n

5 of ceftriaxone, 125 mg, with a single oral dose of azithromycin, 1 g. For prophylaxis against

14 w significantly faster when administered a normal dose of GH than before treatment. Treatment pa

d. psychiatry

adj N of n

12 0-, and 648-mg rectal suppositories. The standard dose of chloral hydrate is 500 to 2,000 mg at

26 arrangements. You will be given a small dose of medication that will make you sleepy.

e. surgery

adj N of n

2 patients allergic to penicillin. Only a single dose of antibiotic is necessary. The placement of

13 than the insulin response to an identical dose of glucose given intravenously (IV), even tho

16. abuse

a. internal medicine

The patterns **N treatment**, **substance N treatment**, and **prep n N** are the particular patterns revealed by the subcorpus of internal medicine:

N treatment

1 psychiatric problems. The major goals of drug abuse treatment are detoxification, abstinence ini

4 or optimism. Physicians can refer to substance abuse treatment specialists; 12-step or mutual-hel

substance N treatment

1 of psychosocial treatment is the backbone of substance abuse treatment, be it psychotherapy, be

prep n N

1 psychiatric problems. The major goals of drug abuse treatment are detoxification, abstinence ini

b. obstetrics and gynecology

prep substance N

11 Nonjudgmental questioning may elicit a history of substance abuse. Pregnancy after Age 35 Current

prep n N

3 ected individuals often have members with alcohol abuse and anxiety disorders. Provocative condition

c. pediatrics

a drug of N

3 direct physiologic effects of a drug of abuse, a medication, or a general medical condition

6 a substance (e.g. a drug of abuse, a medication, or other treatment) or a

d. psychiatry

The subcorpus of psychiatry reveals the patterns **sexual N of adult**, **sexual N of child**, **sexual N of children**, **physical N of adult**, **physical N of child**, **substance ADJ treatment**, and **prep n N**:

sexual N of adult

15 child, physical abuse of adult, and sexual abuse of adult) that frequently are the focus

sexual N of child

25 -TR) lists physical abuse of child, sexual abuse of child, and neglect of child. The

29 physical abuse of child, or for sexual abuse of child). Federal Law the Child Abuse

sexual N of children

43 . Data are from Finklehor D. The sexual abuse of children: current research reviewed. Psyche

45 if the events are disclosed. Table 32-2 sexual abuse of children reported cases in the United

physical N of adult

5 abuse of child, neglect of child, physical abuse of adult, and sexual abuse of adult)

16 of clinical attention is child neglect. Physical abuse of adult this category should be used when

physical N of child

28 correlated with other manifestations of physical abuse of child. (Courtesy of Vincent J. Fontana,

36 abuse or neglect includes five problems (physical abuse of child, sexual abuse of child, neglect

substance ADJ treatment

1 16-year-old male, was admitted to substance abuse treatment for the second time, following a

3 ance Use Among adolescents enrolled in substance abuse treatment programs, 96 percent are polydrug

prep n N

34 be viewed as a form of child abuse, a pedophilia, or a variant of rape.

44 ime prevalence of 17 percent; followed by alcohol abuse, affecting 13 percent; and phobias, affectin

e. surgery

prep n N

11 anaerobic and may be associated with alcohol abuse, foreign body aspiration, a debilitated or e

3 diabetes, hypertension, or a history of tobacco abuse have compromised microvascular circulation t

17. maternal

a. internal medicine

The subcorpus of internal medicine reveals the following patterns for the adjective *maternal*:

prep ADJ serum

3 measurement in the first trimester, followed by maternal serum measurements of MSAFP, HCG, uE3, an

5 dotropin (HCG), and unconjugated estriol (uE3) in maternal serum (triple screen) allowed detection o

prep ADJ n

4 represents one of the leading causes of maternal death in the United States. It occurs

b. obstetrics and gynecology

prep ADJ serum

17 leak into the amnionic fluid, resulting in maternal serum AFP levels that may be dramatically

49 cent (Guschmann and associates, 2003). Levels of maternal serum alpha-fetoprotein (MSAFP) may be el

prep ADJ n

19 the fetal heart rate following correction of maternal acidemia. Gestational age influence

32 al: Fetal thyroid-stimulating hormone response to maternal administration of thyrotropin-releasing h

prep the ADJ abdomen

4 For example, therapeutic radiation doses to the maternal abdomen are contraindicated because of a

prep ADJ adj

15 ch with contractions, (3) bearing-down efforts of maternal abdominal muscles, and (4) extension and

c. pediatrics

prep ADJ n

43 are higher in cord blood than in maternal blood because of active placental transfe

d. psychiatry

prep ADJ n

1 creased risk for chromosomal anomalies because of maternal age, and an assessment of recurrence risk

17 outside the home, the traditional rationale for maternal custody has less force today than it

e. surgery

prep ADJ adj

4 ion of the uterus. Figure 45-14 Location of maternal normal appendix during fetal gestation.

18. psychiatric

a. internal medicine

The subcorpus of internal medicine reveals the pattern **ADJ conditions v** for the adjective *psychiatric*:

ADJ conditions v

3 n, acute nonorganic psychotic disorders, or other psychiatric conditions is indicated.

b. obstetrics and gynecology

n of ADJ disorders

5 (2003) reported a 14-percent point prevalence of psychiatric disorders during pregnancy. Unfortunate

pl-n of ADJ disorders

1 There are only a few reports of psychiatric disorders and pregnancy outcomes. In a

c. pediatrics

n of ADJ disorders

14 to thrive, rash, and hypoglycemia. Simulation of psychiatric disorders occurs, but is less frequent

ADJ pl-n

13 . The clinician must rule out medical and psychiatric conditions that could mimic these symp

d. psychiatry

The patterns **n prep ADJ disorders**, **pl-n prep ADJ disorders**, **prep ADJ symptoms prep**, and **prep ADJ conditions prep** are revealed by the subcorpus of psychiatry:

n prep ADJ disorders

15 such as music. Most common hallucination in psychiatric disorders. Aura (1) Warning sensations

pl-n prep ADJ disorders

2 gies to discover candidate vulnerability genes in psychiatric disorders. Am J Psychiatry. 2003; 160:6

22 contrast to the insomnia in patients with psychiatric disorders, daytime adaptation is gener

prep ADJ symptoms prep

30 setting. Because of the high risk for psychiatric symptoms in abused and neglected child

41 Health Questionnaire (GHQ) Medical patients with psychiatric symptoms of anxiety or depression Self

prep ADJ conditions prep

17 spectrum of diseases and disorders, ranging from psychiatric conditions to cancer, osteoporosis, an

e. surgery

ADJ pl-n

4 vascular necrosis of the femoral head, cataracts, psychiatric problems, osteoporosis, and weight gai

6 yroidectomy, as does respiratory muscle capacity. Psychiatric symptoms such as mental dullness, conf

19. acid

a. internal medicine

The subcorpus of internal medicine reveals the particular patterns **N base balance**, **N base disorders**, **N base disturbance**, **N base disturbances**, **prep n N**, and **n N n**:

prep n N

28 tastatic carcinoma. Elevated levels of hyaluronic acid and special stains and electron microscopy of

N base balance

6 , and alkalemic, but because Na+, K+, and acid-base balance are intrinsically linked, life-t

8 or alkalemia. It is customary to define acid-base balance in terms of the hydrogen

N base disorders

22 ing hypernatremia or hyponatremia, hypercalcemia, acid-base disorders, hypoglycemia and hyperglycemi

23 two values may help identify more complicated acid-base disorders. Much larger increases in the

N base disturbance

27 ble primary cause, such as electrolyte imbalance, acid-base disturbance, hypoxemia, drug toxicity, a

30 body's homeostatic mechanisms to correct the acid-base disturbance. It is often recommended tha

N base disturbances

32 . The mechanisms for changes in potassium in acid-base disturbances are not completely understo

33 is not required for there to be acid-base disturbances. Because pH is related to

n N n

45 on: an aggressive approach. ASA = acetylsalicylic acid (aspirin); CABG = coronary artery bypass graf

b. obstetrics and gynecology

prep n N

20 grain products to require addition of folic acid. 61 Federal Register 8781, 1996 Food and

22 plasma levels have been described for ascorbic acid, folic acid, vitamin B6 (pyridoxine), and vitamin

n N concentrations

1 blood as hemoglobin and plasma proteins. Amino acid concentrations are higher in the fetal than

Lysergic N diethylamide (LSD)

1 Classically known as lysergic acid diethylamide (LSD), these are amine alkaloids obtained only thro

c. pediatrics

prep n N

23 avoided. Drying the auditory canals with acetic acid (2%), Burow solution, or diluted isopropyl al

n N n

22 llowed up by performing quantitative plasma amino acid analysis, measuring phenylalanine and tyrosin

50 osalicylate drugs, which deliver 5-aminosalicylic acid (5-ASA) to the distal gut. Because it

N base balance

2 include serum BUN, creatinine, electrolytes, and acid-base balance. A serum uric acid may

4 in pH can be fatal. Control of acid-base balance depends on the kidneys, the

N base disorders

4 alkalosis (Table 37-5). Treatment of respiratory acid-base disorders focuses on correction of the

5 spiratory compensation for the six primary simple acid-base disorders (Table 37-1). The appropriate

N base disturbance

1 when there is more than one primary acid-base disturbance. An infant with bronchopulmo

2 Diabetic ketoacidosis without specific acid-base disturbance Tubular toxins (amphot

d. psychiatry

n N n

1 ral: tablets and combination with acetylsalicylic acid, acetaminophen, liquid Oxycodonec 15 oral30 o

30 xor) a Mood stabilizers Lithium (Eskalith) Valproic acid (Depakene) Carbamazepine (Tegretol) Topiramat

prep N n

20 ovarian cancers. Autoantibodies to -aminobutyric acid (GABA)-ergic neurons in the serum and

n N concentrations

20 by the positive correlation between homovanillic acid concentrations in the cerebrospinal fluid and

21 toin concentrations up to 25 percent and valproic acid concentrations 11 percent; it does not affect

Lysergic N diethylamide (LSD)

39 least two major substances of abuse, lysergic acid diethylamide (LSD) and 3, 4-methylenedioxymeth

45 Hallucinogens The hallucinogens include lysergic acid diethylamide (LSD), phencyclidine (PCP), and psil

e. surgery

n N n

24 nes (CCK hepatobiliary 2, 6-dimethyl-iminodiacetic acid [HIDA]). These agents are iminodiacetic acid

N base balance

1 patient's vital signs and restoration of acid-base balance can occur. Deciding when the

n N concentrations

1 gastric fluid can be analyzed for bile acid concentrations. On endoscopy, the mucosa is f

20. protein

a. internal medicine

The noun *protein* reveals the patterns **activated protein C**, and **N and n** in the subcorpus of internal medicine:

activated protein C

3 sis of plasminogen activator inhibitor-1. Activated protein C also decreases apoptosis, leuk

24 patients with a high risk of death, activated protein C therapy also improves surviva

N and n

6 to inaccuracies. A spot urine sample for protein and creatinine can be used to estimate

23 1200 mg/day of calcium and reduced animal protein and sodium. [1] The available data support

b. **obstetrics and gynecology**

activated protein C

2 The most widely studied include resistance to activated protein C (aPC) caused by the factor

N and n

3	red nearly ideal sources of nutrients, especially	protein and calcium, for pregnant or lactating wom
10	inhibitory protein, or liver-enriched activatory	protein and PR-A in human endometrial stromal

c. pediatrics

N and n

23	in signal transduction, including B cell linker	protein and Btk. At the pre-B cell
25	n the classification of pediatric undernutrition.	Protein and calorie malnutrition may be associated

d. psychiatry

percent N bound

6	affect its absorption, and it is 55 percent	protein bound in the plasma; 94 percent of lamotri
9	tiagabine is 95 percent, and it is 96 percent	protein bound. It has a half-life of 7

N and n

3	concentration. This initiates a cascade of protein–	protein and protein–lipid interactions in which ne

e. surgery

activated protein C

5	tte smoking, hypertension, and surgery. Activated	protein C is an extremely potent anticoagulant. Re
6	ely potent anticoagulant. Resistance to activated	protein C may be transmitted as an autosomal

N and n

2	the liver some of the demands of	protein and amino acid metabolism. [14] The liver
11	either transudates or exudates based on fluid	protein and lactate dehydrogenase (LDH) concentrate

Table 6.1 summarises the above findings.

Table 6.1: The particular patterns of the 20 top words of the MAWLOT

No.	Word	Internal medicine	Obstetrics & Gynecology	Pediatrics	Psychiatry	surgery
1	**PATIENTS**					
	in PL-N with n	18	16	15	11	18
	PL-N with n	27	17	35	29	32
2	**DISORDER**					
	adj N	5	18	11	12	11
	N and	5	6	7		2
	N is characterized by	5	-	5	15	1
	N is associated with	-	-	-	7	-
	N is pp	11	-	15	-	1
	N is n	-	-	18	5	-
	N is adj	-	-	10	23	-
3	**DIAGNOSIS**					
	N of acute MI	8	-	-	-	-

No.	Word	Internal medicine	Obstetrics & Gynecology	Pediatrics	Psychiatry	surgery
	N of acute n	14	-	8	-	12
	N of n	27	-	31	18	22
	N of adj	-	16	12	17	12
	N of acute	-	-	-	-	8
	N of acute abdominal pain	-	-	-	-	4
	N of acute appendicitis	-	-	-	-	5
	N of adrenal insufficiency	-	-	-	-	7
	N is based on	-	1	11	-	7
	N is confirmed by	-	2	13	-	3
	N is established	-	-	6	1	7
	N is made	2	15	4	6	23
	N is adj	-	-	7	-	-
	N is made by	-	2	7	-	16
4	**THERAPY**					
	N for acute	11	1	2	-	-
	N for adj	13	12	4	12	11
	N for n	18	34	20	13	26
	N is pp	32	27	15	12	21
	N is adj	-	7	5	17	11
	N is n	-	2	4	15	1
5	**SYNDROME**					
	N is n	13	2	12	3	9
	N is pp	-	2	-	12	6
	N is pp pre	14	2	1	9	-
	N is associated with	-	2	6	-	2
	N is caused by	-	5	4	-	1
	N is characterized by	-	5	15	6	3
	N is adj	-	4	4	6	10
6	**FETAL**					
	prep cells of ADJ origin	5	-	-	-	-
	ADJ heart n	-	27	-	-	
	ADJ heart rate	-	39	3	-	-
	ADJ heart rate abnormalities	-	7	1	-	-
	ADJ heart rate acceleration	-	10	-	-	-
	ADJ alcohol syndrome	-	7	17	6	-
	during ADJ development	-	1	-	-	4
	during ADJ life	-	2	4	4	4
7	**CELL**					
	T N activation	11	-	2	-	-
	B N activation	4	-	-	-	-
	sickle N disease	-	7	38	-	-
	sickle N anemia	10	5	18	-	6
	sickle N n	18	8	-	-	1
	islet N tumors	-	-	-	-	15
	germ N tumors	1	5	18	-	16
	squamous N carcinomas of the n	1	-	-	-	7

No.	Word	Internal medicine	Obstetrics & Gynecology	Pediatrics	Psychiatry	surgery
	squamous N carcinomas	4	-	-	-	16
	n N n	-	-	10	8	13
	adj N carcinomas	-	-	-	-	14
8	**CLINICAL**					
	adj ADJ	9	2	3	2	-
	ADJ and adj	21	17	4	5	17
	ADJ manifestations	36	30	35	-	30
	ADJ manifestations prep	3	8	-	-	1
	ADJ manifestations of	27	3	-	-	12
9	**ACUTE**					
	ADJ adj	9	5	7	4	-
	an ADJ abdomen	-	-	1	-	16
	prep ADJ abdomen	-	-	-	-	9
	prep ADJ abdominal pain	-	-	7	-	24
	ADJ stress disorder	-	-	-	23	-
	ADJ stress disorder, PTSD	-	-	-	5	-
	ADJ respiratory acidosis	5	-	-	-	-
	ADJ respiratory alkalosis	5	-	-	-	-
	ADJ respiratory distress syndrome	41	-	9	-	-
	ADJ renal failure	45	-	-	-	-
10	**INFANT**					
	N mortality	2	14	7	-	-
	N mortality rate	-	4	1	-	-
	adj N	2	15	8	5	-
	N is pp	-	2	11	4	1
	the N is -ing	-	-	8	1	-
	the N is adj	-	-	10	-	-
11	**DRUG**					
	N interactions	7	4	3	50	-
	N is pp prep	24	12	3	-	-
12	**CANCER**					
	prep n N in pl-n	10	-	-	-	-
	N is adj	-	-	1	1	-
	N is adj n	-	-	-	-	13
	N is pp	-	1	1	-	18
	N is pp prep	1	3	-	-	19
	N in n	2	9	-	-	15
	in N patients	-	-	-	-	6
	N in patients with	-	-	-	-	7
13	**ADOLESCENT**					
	ADJ n	2	3	7	10	-
	ADJ pl-n	-	-	11	3	3
	ADJ pregnancy	-	3	3	-	-
	ADJ boys	-	-	6	5	-
	ADJ girls v	-	-	-	3	-

No.	Word	Internal medicine	Obstetrics & Gynecology	Pediatrics	Psychiatry	surgery
	or ADJ antisocial behaviour	-	-	-	5	-
14	**CHRONIC**					
	acute or ADJ	15	2	13	2	4
	prep acute or ADJ	6	-	1	-	10
	prep ADJ obstructive lung disease	5	-	-	-	-
	ADJ obstructive pulmonary disease	17	-	1	-	-
	prep ADJ obstructive pulmonary disease	18	-	-	-	-
	prep ADJ kidney disease	30	-	-	-	-
	prep ADJ renal failure	-	-	14	-	10
	prep ADJ fatigue syndrome	-	-	30	24	-
	prep ADJ pancreatitis	-	-	3	-	31
15	**DOSE**					
	adj N of n	18	9	27	27	7
16	**ABUSE**					
	N treatment	4	-	-	3	-
	prep n N	1	3	-	4	7
	Substance N treatment	1	-	-	3	-
	a drug of N	-	-	7	-	-
	sexual N of adult	-	-	-	10	-
	sexual N of child	-	-	-	8	-
	sexual N of children	-	-	-	3	-
	physical N of adult	-	-	-	6	-
	physical N of child	-	-	-	9	-
17	**MATERNAL**					-
	prep ADJ n	8	8	3	16	-
	prep ADJ adj	-	2	-		1
	prep the ADJ abdomen	-	10	-	-	-
	prep ADJ serum	5	11	-	-	-
18	**PSYCHIATRIC**					
	prep ADJ conditions prep	-	-	-	4	-
	ADJ conditions v	3	-	-	1	-
	ADJ pl-n	-	-	3	-	4
	n of ADJ disorders	-	2	1	9	-
	pl-n prep ADJ disorders	-	1	-	8	-
	prep ADJ symptoms prep	-	-	-	9	-
19	**ACID**					
	N base balance	11	1	4	-	1
	n N concentrations	-	3	-	3	1
	Lysergic N diethylamide (LSD)	-	1	-	11	-
	n N n	6	-	2	24	25
	prep n N	13	11	5	11	
	N base disorders	4	-	5	-	-
	N base disturbance	7	-	2	-	-

No.	Word	Internal medicine	Obstetrics & Gynecology	Pediatrics	Psychiatry	surgery
	N base disturbances	11	-	-	-	-
20	**PROTEIN**					
	activated N C	24	3	-	-	7
	percent N bound	-	-	5	5	-
	N and n	7	6	30	-	9

6.4 Frequent collocation/phraseological patterns of medical technical words of the MAWLOT

To find the most frequent patterns of the MAWLOT's medical technical words, 4 statistically top words (0.010 % - 0.015 %) out of 53 words (see Table 5.3) were selected to be considered for finding their frequent patterns. These words are *cesarean* (F= 529, R= 4), *esophageal* (F=512, R=5), *etiology* (F=478, R=5), and *anemia* (F=385, R=5).

1. cesarean

a. internal medicine

There is no discerning pattern similar as the following ones in the subcorpus of internal medicine for the noun *cesarean*.

b. obstetrics and gynecology

The patterns **number of N deliveries**, **risk of N delivery**, **rate of N delivery**, **vaginal birth after N delivery**, and **in women with a prior N delivery** are revealed by the subcorpus of obstetrics and gynecology:

number of N deliveries

4 ized labor management protocol reduced the number of cesarean deliveries for dystocia. Their overall

6 e or more changes 30 percent increased number of cesarean deliveries 26 percent stopped perf

risk of N delivery

12 pidural analgesia prolong labor and increase risk of cesarean delivery? A natural experiment. Am J

50 duction with cervical ripening increases the risk of cesarean delivery in multiparous women. Obstet

rate of N delivery

13 years between 1989 and 1996 when the annual rate of cesarean delivery actually decreased (Fig. 25-1

27 , Hamar and associates (2001) found that the rate of cesarean delivery following elective induction

vaginal birth after N delivery

29 necologists: Induction of labor for vaginal birth after cesarean delivery. Committee Opinion No. 271

in women with a prior N delivery

23 50 percent of placentas in women with a prior cesarean delivery had adhered myometrial fib

28 prostaglandin E1 use in women with a prior cesarean delivery. Indeed, Plaut and associa

c. pediatrics

The subcorpus of pediatrics reveals the patterns **delivery by N section**, **prep N section**, and **N section**:

delivery by N section

2 infant in cases of NATP and ITP. Delivery by cesarean section is recommended to pre

3 second born of twins, and delivery by cesarean section without labor. RDS may develop

prep N section

3 all newborns, whether delivered vaginally or by cesarean section, as soon as possible after deliver

7 small infants and an increased requirement for cesarean section. If tumor extends into the renal

N section

4 ly Treat maternal genitourinary lesions; possible cesarean section? FFP, fresh frozen plasma

12 or Presentation-vertex, breech Vaginal versus cesarean section Spontaneous labor versus augmen

d. psychiatry

N section

1 support during labor reduces the rate of cesarean section and forceps deliveries, the need

e. surgery

There is no discerning pattern similar as the above ones.

2. esophageal

a. internal medicine

The subcorpus of internal medicine reveals the pattern **ADJ reflux**:

ADJ reflux

1 individuals who are suspected of having recurrent esophageal reflux and aspiration events at night.

4 occur in normal persons during sleep; also, esophageal reflux causing cough and asthma symptom

b. obstetrics and gynecology

The pattern **ADJ atresia** is revealed by the subcorpus of obstetrics and gynecology:

ADJ atresia

1 cts, anal atresia, tracheoesophageal fistula with esophageal atresia, and radial dysplasia. It is re

2 with a number of abnormalities. These include esophageal atresia, diaphragmatic hernia, and abdominal

c. pediatrics

The subcorpus of pediatrics reveals the patterns **ADJ varices** and **ADJ atresia** for the adjective *esophageal*:

ADJ atresia

1 invariably present in children who have had esophageal atresia and a tracheoesophageal fistula

5 is often present at birth. Infants with esophageal atresia have a history of polyhydramnio

ADJ varices

3 atresia. After this age, peptic disease and esophageal varices continue to be common causes of

9 NSAID)-induced gastric injury causes hematemesis; esophageal varices occur in this age group with

d. psychiatry

adj ADJ pl-n

4 Functional Gastrointestinal Disorders Functional esophageal disorders Globus Lump in throat, common

14 functional esophageal disorder other nonspecific esophageal symptoms Functional gastroduodenal diso

e. surgery

The patterns of **n prep ADJ cancer**, **ADJ motility disorders**, **n prep ADJ varices**, and **n of the ADJ body** are the most frequently patterns revealed by the subcorpus of surgery:

n prep ADJ cancer

3 nting with esophageal cancer. The distribution of esophageal cancer across gender, age, and race is

17 as the cornerstone of combination therapy in esophageal cancer. As a single agent, it has

ADJ motility disorders

8 secondary motor disorders of the esophagus. Most esophageal motility disorders fall into one of fiv

10 diagnosis and will help eliminate other potential esophageal motility disorders. In typical achalasi

n prep ADJ varices

5 varices most commonly occur in association with esophageal varices, but they occasionally occur al

n of the ADJ body

2 of information about the function of the esophageal body and the LES may be obtained

3 LES, as well as dilation of the esophageal body, are observed. A lack of peristalt

3. etiology

a. internal medicine

The noun *etiology* reveals the pattern **the N prep n**, which means that the noun is followed by the specific preposition *of* and a noun:

the N of n

2 and physical examination often help reveal the etiology of AKI in a patient. The history

11 to the prostate gland. Theories for the etiology of type III prostatitis include infectious

b. obstetrics and gynecology

The particular pattern revealed by the subcorpus is **the N of n**, which means that **N** is preceded by the article *the*, and followed by the specific preposition *of* and a noun:

the N of n

4 hologically enlarged thyroid gland depends on the etiology of hypothyroidism and is more likely in

18 viously unscarred uterus. A classification of the etiology of uterine rupture is presented in Table u

c. pediatrics

The subcorpus reveals the patterns **the N of n**, and **the N of adj**:

the N of n

10 system fails, the apneic child dies. The etiology of apnea is diverse and not well

29 infusions of platelets, blood, or both. The etiology of DIC in a newborn includes hypoxia,

the N of adj

32 of the world, iodine deficiency is the etiology of endemic goiter (endemic cretinism). Th

37 to approximately one per day. Etiology The etiology of functional constipation and soiling in

d. psychiatry

adj N

6 syndrome is an illness with a controversial etiology and pathogenesis. Besides persistent fati

15 is a connective tissue disease of unclear etiology, characterized by recurrent episodes of d

e. surgery

The subcorpus of surgery reveals the particular patterns **the N of n**, **the N of adj**, and **adj N**:

the N of n

10 the specific population being staged and the etiology of HCC in that particular population. The

12 of the disease. Another theory of the etiology of IBD concerns an altered immunologic re

adj N

14 careful bleeding history may suggest a specific etiology, consideration must always be given to pr

the N of adj

19 5 Pleural fluid LDH 1.67 times normal serum The etiology of pleural effusions is quite varied.[29]

20 sinus nerve have been implicated in the etiology of postoperative hypertension and hypoten

4. anemia

a. internal medicine

The patterns **N and n**, **N is adj**, and **N with n** are revealed by the subcorpus of internal medicine:

N and n

2 conditions that limit oxygen delivery, such as anemia and carbon monoxide intoxication. Angina or

12 a complete blood cell count to exclude anemia and thyroid function tests to exclude hyper

N is adj

3 Transfusion of Erythrocytes and Erythropoietin Anemia is common in septic shock, but the

N with n

1 . Early treatment of acidosis with oral citrate, anemia with erythropoietin, and secondary hyperpar

2 achieve correction of metabolic abnormalities and anemia with restoration of normal renal function.

b. obstetrics and gynecology

N and n

3 shunting within the placenta, leading to fetal anemia and hydrops (see Chap. 29, Nonimmune Hydrop

15 9-3). Fetal hepatic abnormalities are followed by anemia and thrombocytopenia, then ascites and hydr

c. pediatrics

The subcorpus of pediatrics reveals the particular patterns **N is adj**, **N and n,** and **N with n**:

N and n

12 nes leads to hydrops fetalis, severe intrauterine anemia and death, unless intrauterine transfusion

23 and thalassemia. Parvovirus B19 also causes fetal anemia and hydrops fetalis after primary infection

N is adj

44 and ineffective erythropoiesis in the marrow. The anemia is severe and leads to growth failure

50 simple transfusion therapy is indicated when the anemia is symptomatic. Vaso-occlusive crises may

N with n

3 be discovered incidentally to a severe hemolytic anemia with growth failure, splenomegaly, and chro

6 ?+? + 20-40 5 60-80 severe hypo/micro anemia with Hb 7-9 g/dL, hepatosplenomegaly, bone

d. psychiatry

The subcorpus of psychiatry reveals the pattern **N and n**, which means that the noun is followed by the conjunction *and* and a noun:

N and n

2 ic workup Dementia Part of workup of megaloblastic anemia and dementia B12 deficiency associated with

6 first step in treatment is looking for anemia and treating it, if found. Benzodiazepines

e. surgery

The subcorpus of surgery reveals the particular pattern **N is adj**:

N is adj

4 ts and improvement in physical activity. Finally, anemia is common in uremic patients. It is

8 toilet bowl. Chronic occult bleeding leading to anemia is rare, and other causes of anemia

Table 6.2 summarises the frequent particular patterns of the selected medical technical words of the Medical Academic Word List of Textbooks (MAWLOT) revealed by the Medical Academic Corpus of the study.

Table 6.2: The particular patterns of the medical technical words of the MAWLOT

No.	Word	Internal medicine	Obstetrics & Gynecology	Pediatrics	Psychiatry	surgery
1	**CESAREAN**					
	delivery by N section	-	16	3	-	-
	N section			5	1	-
	prep N section			13	-	-
	number of N delivery	-	-	-	-	-
	risk of N delivery	-	15	-	-	-
	rate of N delivery	-	5	-	-	-
	vaginal birth after N delivery	-	17	-	-	-
	in women with a prior to N delivery	-	10	-	-	-
2	**ESOPHAGEAL**					
	ADJ reflux	10	-		-	-
	ADJ atresia	-	8	12	-	-
	ADJ varices	-	-	13	-	-

	adj ADJ pl-n	-	-	-	4	-
	n prep ADJ cancer	-	-	-	-	22
	ADJ motility disorders	-	-	-	-	6
	N prep ADJ varices	-	-	-	-	8
	N of the ADJ body	-	-	-	-	11
3	**ETIOLOGY**					
	the N of n	4	16	13	-	8
	the N of adj	-	-	8	-	4
	adj N	-	-		9	7
4	**ANEMIA**					
	N is adj	1	-	6	-	10
	N and n	9	17	1	6	-
	N with n	3	-	5	-	-

6.5 Discussion on the resulted patterns

To investigate the differences and similarities in frequency distribution of patterns of the words included in MAWLOT, 5 nouns and 2 adjectives were selected to consider and compare from the list of the top 20 words (see section 6.3). The list contains 14 nouns and 6 adjectives. The selected nouns are *patients* **(PL-N)**, *protein* **(N)**, *diagnosis* **(N)**, *dose* **(N)**, and *cell* **(N)**. The selected adjectives are *acute* and *maternal*.

Patients **(PL-N)**, as the most frequent word in the list, is considered versus *protein* **(N)**, as the least frequent word in the list. *Diagnosis* **(N)**, as the word which presents the most different types of patterns (14 types of patterns) in the list, is considered versus *dose* **(N)**, as the word which presents the least different types of patterns (just 1 type of pattern). Also, the noun *cell* **(N)** is considered as the most frequent word in the list of Wang's et al. (MAWL) (the seventh word in the MAWLOT). Two adjectives, i.e. *acute* and *maternal* are also considered and compared. The word *acute* **(ADJ)** is considered as the adjective, which reveals the most frequent different types of patterns (10 types) versus *maternal* **(ADJ)**, which reveals the least different types of patterns (4 types) in the list.

The first word to consider is the plural noun *patients* **(PL-N)**, which is also the top word in the list of the top 20 words of the MAWLOT (F= 8943, %= 0.25). The selected concordance lines of the corpus of the study reveal two different types of patterns for *patients*: **PL-N with n** and **in PL-N with n**. In the pattern **PL-N with n**, the **PL-N** is followed by the specific preposition *with* and a noun. The pattern **in PL-N with n** reveals that the **PL-N** is preceded by the specific preposition *in*, and followed by the specific preposition *with* and a noun. Considering the frequency distribution of the pattern **in PL-N with n** across the subcorpora it appears that in the subcorpora of internal medicine and surgery the frequency distribution of the pattern is the same (18 times) in the selected concordance lines of the corpus of the study. The frequency distribution is also approximately similar in the subcorpora of obstetrics and gynecology (16 times) and pediatrics (15 times). However, in the subcorpus of psychiatry it is slightly different (11 times). The pattern is totally revealed 78 times by the selected concordance lines of the corpus. The pattern **PL-N with n** is also revealed approximately similar in the subcorpora of pediatrics (35 times) and surgery (32 times). It is also revealed similarly by the subcorpora of internal medicine (27 times) and psychiatry (29 times) in the selected concordance lines of the corpus of the study. The pattern occurs 17 times in the subcorpus of obstetrics and gynecology. In comparison, the pattern **PL-N with n** occurs more frequently (140 times) than the pattern **in PL-N with n** (78 times) in the Medical Academic Corpus of the study.

The word *protein* (**N**), as the least frequent word in the list of the top 20 words of the MAWLOT (F=1396, %= 0.04), reveals 3 different types of patterns: **activated N C**, **percent N bound**, and **N and n**. The pattern **activated N C** reveals that **N** is preceded by the specific past participle *activated*, and followed by the specific noun *C*. The pattern occurs differently across the subcorpora of the study. That is, it occurs 24 times in the subcorpus of internal medicine, but it occurs just 3 times in the subcorpus of

obstetrics and gynecology. It also occurs 7 times in the subcorpus of surgery. The pattern is not revealed by the subcorpora of psychiatry and surgery. In total, it occurs 34 times in the selected lines of the corpus. The pattern **percent N bound** reveals that **N** is preceded by a noun and followed by a noun as well. The specific noun to precede **N** is *percent*, and the specific noun to follow the **N** is *bound*. The pattern occurs similarly in the subcorpora of pediatrics and psychiatry (5 times), but it is not revealed by the subcorpora of internal medicine, obstetrics and gynecology, and surgery. The pattern occurs totally 10 times in the selected concordance lines of the corpus. The last particular pattern of the noun *protein* is **N and n**, which means that the **N** is followed by the conjunction *and* and a noun. The pattern occurs approximately similar in the subcorpora of internal medicine (7 times), obstetrics and gynecology (6 times), and surgery (9 times). It occurs 30 times in the subcorpus of pediatrics, but it does not occur in the subcorpus of psychiatry. In total, it occurs 52 times in the selected concordance lines of the corpus.

In comparison, the types of patterns revealed by the selected concordance lines of the corpus for the word *patients* is less (2 types) than the revealed different types of patterns for the word *protein* (3 types). On the other hand, the different types of *patients'* patterns occur in all 5 subcorpora, but the *protein*'s patterns does not occur in all subcorpora (they occur in 2, 3, and 4 subcorpora). Also, the frequency of the revealed patterns for *patients* is more (218 times) than the frequency of the revealed patterns for the word *protein* (96 times) in the selected concordance lines of the corpus. In addition, the *patients*' patterns are approximately similar in their structure, i.e. both patterns **(PL-N with n** and **in PL-N with n)** are followed by the specific preposition *with* and a noun. Finally, considering the patterns it clears that the word *patients* tends to be combined with the specific preposition *with* and a noun, while the tendency of *protein* is mostly

toward the combination with a noun. Table 6.3 shows the words with their particular patterns.

Table 6. 3: The particular patterns of the words *patients* and *protein*

patients	**PL-N with n** **in PL-N with n**
protein	**activated N C** **percent N bound** **N and n**

The other two words to consider and compare are *diagnosis* **(N)** and *dose* **(N)**. The word *diagnosis* is considered as the word, which reveals the most different types of patterns (14 types) in the list of the top 20 words of the MAWLOT. In contrast, the word *dose* reveals just 1 type of pattern in the corpus. The 14 different types of the patterns of *diagnosis* can be divided in two groups. The first group are the patterns, which **N** is followed by the specific preposition *of*, i.e. **N of** (8 types). The second group are the patterns, which **N** is followed by the specific verb *is*, i.e. **N is** (6 types).

The first group of patterns of the word *diagnosis*, i.e. **(N of)** are: **N of acute MI**, **N of acute n**, **N of n**, **N of adj**, **N of acute**, **N of acute abdominal pain**, **N of acute appendicitis**, and **N of adrenal insufficiency**. The pattern **N of acute MI** shows that the **N** is followed by the specific preposition *of*, the specific adjective *acute*, and the specific noun *MI* (the abbreviation for myocardial infarction). This pattern is just revealed 8 times by the subcorpus of internal medicine, and is not revealed by other subcorpora. The pattern **N of acute n** occurs approximately similar in the subcorpora of internal medicine (14 times) and surgery (12 times). In the subcorpus of pediatrics the pattern occurs 8 times, but it is not revealed by the subcorpus of obstetrics and

gynecology. In total, the pattern occurs 34 times in the selected concordance lines of the corpus. The pattern **N of n** occurs approximately similar in the subcorpora of internal medicine (27 times) and surgery (22 times). In the subcorpus of pediatrics it occurs 31 times, but it does not occur in the subcorpus of obstetrics and gynecology. The pattern occurs 18 times in the subcorpus of psychiatry. In total, the pattern occurs 98 times in the selected concordance lines of the corpus of the study. Finally, the patterns **N of acute**, **N of acute abdominal pain**, and **N of adrenal insufficiency** are revealed just by the subcorpus of surgery.

The second group of patterns of the word *diagnosis* **(N is)** are: **N is based on**, **N is confirmed by**, **N is established**, **N is made**, **N is adj**, **N is made by**. The pattern **N is based on**, which means that *diagnosis* is followed by the specific verb *is*, the specific past participle *based* and the specific preposition *on*, occurs 11 times in the subcorpus of pediatrics, and 7 times in the subcorpus of surgery. However, it just occurs 1 time in the subcorpus of obstetrics and gynecology. The pattern is not revealed by the subcorpora of internal medicine and psychiatry. The pattern **N is confirmed by** occurs differently in the 5 subcorpora. That is, while the subcorpora of internal medicine and psychiatry do not reveal the pattern, the pattern is revealed approximately similar in the subcorpora of obstetrics and gynecology (2 times), and psychiatry (3 times). In the subcorpus of pediatrics the pattern occurs 13 times. The pattern **N is established** occurs approximately similar in the subcorpora of pediatrics (6 times) and surgery (7 times). The pattern is not revealed by the subcorpora of internal medicine and obstetrics and gynecology. It also occurs just 1 time in the subcorpus of psychiatry. In total, the pattern occurs 14 times in the selected concordance lines of the corpus. The pattern **N is made**, which means that the **N** is followed by the specific verb *is* and the specific past participle *made*, occurs in all 5 subcorpora, but differently. That is, in the subcorpora of pediatrics (4 times) and psychiatry (6 times) it occurs approximately similar, in the

subcorpus of internal medicine 2 times, in the subcorpus of obstetrics and gynecology 15 times, and in the subcorpus of surgery 23 times. In total, the pattern occurs 50 times in the selected concordance lines of the corpus of the study. The pattern **N is made by** also occurs differently in the subcorpora. That is, while it is not revealed by the subcorpora of internal medicine and psychiatry, it occurs 16 times in the subcorpus of surgery. In the subcorpus of obstetrics and gynecology it also occurs just 2 times. In total, the pattern is revealed 25 times by the selected concordance lines of the corpus (half of the pattern **N is made**). Finally, the pattern **N is adj**, which means that **N** is followed by the specific verb *is* and an adjective, occurs just in the subcorpus of pediatrics (7 times).

In conclusion, considering the word *diagnosis* **(N)** in the subcorpus of internal medicine, it appears that **N** tends to be combined with the specific preposition *of* (49 times), while the combination of **N** with the verb *is* occurs just 2 times in the subcorpus. In contrast, in the subcorpus of obstetrics and gynecology, the **N** tends to be combined with the specific verb *is* (20 times). The combination of **N** with the specific preposition *of* occurs 16 times in the subcorpus. The subcorpus of pediatrics reveals similarity in the tendency towards the combination of **N** with *of* and *is***.** That is, in the subcorpus of pediatrics the **N** tends to be combined 51 times with the specific preposition *of*, and 48 times with the specific verb *is*. Also, the behaviour of **N** in the subcorpus of psychiatry is similar to the behaviour of **N** in the subcorpus of internal medicine, in which the **N** tends to be combined with the specific preposition *of* (35 times) in the subcorpus, while the combination of **N** with the specific verb *is* occurs just 7 times. In addition, in the subcorpus of surgery **N** behaves similar to the other subcorpora, i.e. **N** tends to be combined with the specific preposition *of* (70 times). The combination of **N** with the verb *is* occurs 56 times. In conclusion, the word *diagnosis* tends to be combined with

the specific preposition *of (*221 times) in the selected concordance lines of the corpus, while the combination of **N** with the specific verb *is* occurs 133 times.

Furthermore, the word *dose* **(N)** is considered as the word, which reveals the least different types of patterns (1 type) in the list of the top 20 words of the MAWLOT. The particular pattern revealed by the corpus of the study is **adj N of n**, which means that the **N** is preceded by an adjective, and followed by the specific preposition *of* and a noun. The frequency distribution of the pattern across the subcorpora of pediatrics and psychiatry is the same (27 times). It also occurs approximately similar in the subcorpora of obstetrics and gynecology (9 times), and surgery (7 times). The pattern is also revealed 18 times in the subcorpus of internal medicine. In total, the pattern occurs 88 times in the selected concordance lines of the corpus.

As a result, considering the revealed patterns of two words *diagnosis* and *dose* it appears that the tendency of both words towards the combination of **N** with the specific preposition *of* is the same, but the frequency distribution of this combination is different across the subcorpora. Table 6.4 shows the particular patterns of the words *diagnosis* and *dose*.

Table 6.4: The particular patterns of the words *diagnosis* and *dose*.

diagnosis	**N of acute MI**
	N of acute n
	N of n
	N of adj
	N of acute
	N of acute abdominal pain
	N of acute appendicitis

	N of adrenal insufficiency **N is based on** **N is confirmed by** **N is established** **N is made** **N is adj** **N is made by**
dose	**adj N of n**

The last noun to consider is *cell* **(N)**, as the top word in the MAWL's list of Wang et al. (2008) in frequency (F=4421), and the seventh word (F=3428, %=0.09) in the MAWLOT's list of the present study. Considering the revealed patterns by the corpus of the study it shows that *cell* reveals 11 different types of patterns with 255 times of occurrences in the selected concordance lines of the corpus of the study. These patterns are: **T N activation**, **B N activation**, **sickle N disease**, **sickle N anemia**, **sickle N n**, **islet N tumors**, **germ N tumors**, **squamous N carcinomas of the n**, **squamous N carcinomas**, **n N n,** and **adj N carcinomas**. Looking at the patterns it appears that in 10 types of the patterns (out of 11) the **N** is preceded or followed by a noun. The specific nouns to precede **N** are *T*, *B, sickle*, *islet*, *germ*, *squamous*. The specific nouns (singular and plural) to follow the **N** are: *activation*, *disease*, *anemia*, *tumors*, and *carcinomas*. Different combination of **N** with these nouns reveals different diseases. For example, the pattern **sickle N anemia**, which means that the **N** is preceded by the specific noun *sickle*, and followed by the specific noun *anemia*, points to a condition in which there aren't enough healthy red blood cells to carry adequate oxygen throughout body (Brooker, 2008). While, the combination of **N** with the specific noun *squamous* and the specific plural noun *carcinomas* in the pattern **squamous N carcinomas** reveals an

uncontrolled growth of abnormal cells arising in the squamous cells, which compose most of the skin's upper layers (the epidermis) (Brooker, 2008).

Furthermore, the frequency distribution of the above patterns across the 5 subcorpora is also different, and depends on the combination of the pattern with a specific noun. For example, the pattern **T N activation**, **B N activation**, and **sickle N n** occurs with a higher frequency in the subcorpus of internal medicine than the subcorpus of surgery. It is because the diseases specified by the patterns have more clinical treatments (internal medicine) than operational treatments (surgery). In contrast, while the patterns **islet N tumors** and **germ N tumors** are not revealed frequently by the subcorpus of internal medicine (the pattern **germ N tumors** occurs just 1 time in the subcorpus of internal medicine), they occur 15 and 16 times in the subcorpus of surgery. It also occurs because both diseases specified by the patterns are considered as the cases of surgery with an operational treatment. In addition, the pattern **adj N carcinomas**, which means that the **N** is preceded by an adjective, is revealed just by the subcorpus of surgery (14 times). It is also because of the operational treatment of the disease. The pattern is not revealed by other subcorpora.

In conclusion, considering the different types of patterns of the word *cell*, it appears that the **N** tends mostly to be combined (preceded or followed) by a noun. The combination of **N** with different nouns occurs in 10 patterns (out of 11 patterns) with totally occurrences of 241 times in the selected concordance lines of the corpus of the study. The patterns also occur mostly in the subcorpora of pediatrics (86 times) and surgery (88 times). The subcorpus of psychiatry reveals just one pattern **(n N n)** with 8 times occurrences. Table 6.5 shows the particular patterns of the word *cell.*

Table 6.5: The particular patterns of the word *cell.*

cell	**T N activation** **B N activation** **sickle N disease** **sickle N anemia** **sickle N n**, **islet N tumors** **germ N tumors** **squamous N carcinomas of the n** **squamous N carcinomas n N n** **adj N carcinomas**

Furthermore, after considering the frequency distribution of the patterns of 5 nouns (out of 14 nouns in the list of the top 20 words of the MAWLOT) across the subcorpora, two adjectives were considered and compared as well. The list contains 6 adjectives with different types of patterns in the Medical Academic Corpus of the study. The adjective which reveals the most different types of patterns in the list is *acute* with 10 different types of patterns. *Acute* is the ninth frequent word (F=2014, %= 0.05) in the list, which is considered and compared versus the adjective *maternal*. The adjective *maternal* reveals the least different types of patterns (4 types of patterns) in the selected concordance lines of the corpus. *Maternal* is the seventeenth frequent word in the word list of MAWLOT (F= 1543, %= 0.04).

The 10 different types of patterns revealed by the word *acute* **(ADJ)** are: **ADJ adj**, **an ADJ abdomen**, **prep ADJ abdomen**, **prep ADJ abdominal pain**, **ADJ stress disorder**, **ADJ stress disorder PTSD**, **ADJ respiratory acidosis**, **ADJ respiratory alkalosis**, **ADJ respiratory distress syndrome**, and **ADJ renal failure**. The patterns occur totally 215 times in the selected concordance lines of the corpus of the study, but with different frequency distribution across the 5 subcorpora. For example, the pattern **prep ADJ abdomen** occurs 16 times in the subcorpus of surgery, but just 1 time in the

subcorpus of pediatrics. The pattern is not revealed by other subcorpora. Also, the pattern **ADJ stress disorder** occurs 23 times in the subcorpus of psychiatry, but is not revealed by other subcorpora. Also, the pattern **ADJ renal failure** occurs 45 times in the subcorpus of internal medicine, but does not occur in other subcorpora. On the other hand, the pattern **ADJ adj** does not occur in the subcorpus of surgery, but it occurs approximately similar in the subcorpora of internal medicine (9 times) and pediatrics (7 times). It also occurs approximately similar in the subcorpora of obstetrics and gynecology (5 times) and psychiatry (4 times). The patterns **ADJ respiratory distress syndrome** and **ADJ renal failure** occur more frequently (50 and 45 times) in the selected concordance lines of the corpus in comparison with other types of the patterns of **ADJ**. The pattern **ADJ respiratory distress syndrome** is revealed just by the subcorpora of internal medicine and pediatrics. Also, the pattern **ADJ renal failure** is revealed just by the subcorpus of internal medicine.

In conclusion, considering the above different patterns of the adjective *acute* reveals that the **ADJ** tends mostly to be followed by a noun. Just in 3 patterns **(ADJ adj**, **prep ADJ abdominal pain**, and **ADJ renal failure)** the **ADJ** is followed by another adjective. Also, the **ADJ** is preceded just by a preposition or the specific article *an*.

Furthermore, the adjective *maternal* reveals the least different types of patterns (4 types) among 6 adjectives included in the list of the 20 top frequent words. These patterns are: **prep ADJ n**, **prep ADJ adj**, **prep the ADJ abdomen**, and **prep ADJ serum**. The most frequent pattern of the **ADJ** is **prep ADJ n** (35 times), which occurs similarly in the subcorpora of internal medicine and obstetrics and gynecology (8 times), but differently in the subcorpora of pediatrics (3 times) and psychiatry (16 times). The pattern is not revealed by the subcorpus of surgery. The least frequent pattern of **ADJ** is the pattern **prep ADJ adj**, which occurs just 3 times (2 times in

pediatrics; 1 time in psychiatry) in the selected concordance lines of the corpus. The pattern does not occur in the subcorpora of internal medicine, pediatrics, and psychiatry. Two other patterns, i.e. **prep the ADJ abdomen**, and **prep ADJ serum** occur approximately similar in the subcorpus of obstetrics and gynecology (10 and 11 times). The patterns are not revealed by the subcorpora of pediatrics, psychiatry, and surgery. The pattern **prep the ADJ abdomen** is not revealed by the subcorpus of internal medicine, and the pattern **prep ADJ serum** occurs just 5 times in the subcorpus of internal medicine.

As a result, considering the different types of the patterns of the adjective *maternal* it shows that the **ADJ** tends to be preceded just by a preposition, however, it tends to be followed differently by another adjective, noun, or a specific noun (*abdomen*). The frequency distribution of the patterns of the adjective across the subcorpora is also different. That is, the subcorpus of obstetrics and gynecology reveals all types of the patterns of **ADJ**, the subcorpus of internal medicine reveals two types of the patterns (**prep ADJ N, prep ADJ serum**), and the subcorpora of pediatrics and psychiatry reveal just one type of the patterns (**prep ADJ n**). The subcorpus of surgery reveals just one type of the pattern (**prep ADJ adj**) as well. The 4 types of patterns of the adjective *maternal* occur totally 64 times in the selected concordance lines of the corpus of the study. Table 6.6 shows the particular patterns of the words *acute* and *maternal*.

Table 6.6: The particular patterns of the words *acute* and *maternal*

acute	**ADJ adj** **an ADJ abdomen** **prep ADJ abdomen** **prep ADJ abdominal pain** **ADJ stress disorder** **ADJ stress disorder PTSD** **ADJ respiratory acidosis** **ADJ respiratory alkalosis** **ADJ respiratory distress syndrome** **ADJ renal failure**
maternal	**prep ADJ n** **prep ADJ adj** **prep the ADJ abdomen** **prep ADJ serum**

6.6 Summary

In this chapter the results and discussions of the Research Question 2 were presented. This was followed by the frequent collocation/phraseological patterns of the words included in the MAWLOT related to the differences in frequency distributions of the patterns across the subcorpora of the study. The frequent collocation/phraseological patterns of the medical technical words of the study were presented as well.

CHAPTER 7: DISCUSSIONS AND CONCLUSIONS

7.1 Introduction

In this chapter, a summary of the study is first provided. Next, the findings of the study are discussed according to the six research questions. The first three questions deal with the developing a medical academic vocabulary of textbooks (MAWLOT), finding the most frequent collocation/phraseological patterns of the words included in MAWLOT (see Chapter 6), and considering the differences of frequencies of the words across subdisciplines of the corpus of the study. The utility of frequent words and word lists in language teaching, besides the usefulness of collocation/phraseological patterns in language teaching are discussed as well. The other three research questions evaluate the word list resulting from the study. Accordingly, the findings of the descriptive and evaluation phase are discussed respectively. Next, the discussion focuses on the implications of the study, suggestions for application of the word list, and recommendations for further research.

7.2 Summary of the study

This study was conducted in response to the absence of discipline-specific word lists in medical domain that serves the EAP/EMP context. Drawing on and combining methods and procedures from Coxhead (2000) and Wang et al. (2008), the present corpus-based study developed a medical academic vocabulary list of textbooks, the MAWLOT.

The first phase of this study was guided by research questions 1-3, which aimed to identify the MAWLOT's lexical items in terms of frequency and range in the corpus of medical textbooks, in addition to finding the particular patterns of the lexical items and considering the frequency distribution of the patterns of the words included in MAWLOT across the subcorpora of the study. The second phase of the study aimed to

identify the MAWLOT's coverage in comparison with the GSL's and AWL's coverage to prove that the former word list may serve better medical learners' needs.

Identifying exact words in order to ensure that they are worth being studying is the crux of the present study. In developing the word list, 459 words met the criteria for inclusion in the Medical Academic Word List of Textbooks (MAWLOT) (see Appendix D). These words were selected after modifying a primary list of 665 words, which all had met the criteria for inclusion in the word list (see Appendix C for the removed words). In addition, by studying the words in the concordances students are expected to explore the relationships between words by accessing to accurate information about the way language is authentically used. A list of medical technical words was also provided to increase the students' vocabulary knowledge in their reading medical texts (see Table 5.3).

As for data collection, the MAWLOT was developed from the Medical Academic Corpus, a corpus comprising 3.5 million running words of academic textbooks in medicine. The words for inclusion in the word list were selected based on the criteria used by out of the GSL (West, 1953) and the AWL (Coxhead, 2000) (see section 4.3.1), i.e. words which had occurred 100 times or more in the corpus (at least 20 times in each subcorpus). Furthermore, the words had to occur in 3 or more (out of 5) subcorpora. These target words were used to measure their different frequency of occurrences in the subcorpora. To retrieve the corpus information, the Compleat Lexical Tutor (CLT) and the AntConc were selected as the instruments to analyze the data.

In addition, measuring the differences of words' occurrences included in the MAWLOT indicated that the word list is slightly advantageous for the discipline of pediatrics with coverage of 12.51%, which is similar to the coverage of internal medicine (12.14%). The coverage of the word list for surgery is 11.78%, for obstetrics and gynecology 10.74%, and for psychiatry 9.21%. As a result, the overall coverage of

the MAWLOT is 11.27% of the vocabulary in Medical Academic Corpus. Furthermore, the effect of medical technical words on increasing students' vocabulary knowledge in reading medical textbooks was considered. The word list was also examined in similar texts of a different collection to have a real test of the word list (Coxhead, 2000). The findings showed that the MAWLOT's word families occur differently, but with good coverage (10.48%) in the subcorpora of different collection of texts, i.e. the Second Corpus. Finally, the MAWLOT was compared in terms of tokens (running words), types, the number of words included in the word lists, and the coverage with the AWL (Coxhead, 2000) and the MAWL (Wang et al., 2008) to find out how the MAWLOT is an improvement on existing word lists to be used by medical students in reading medical textbooks.

7.3 Discussions and conclusions of the findings

Discussions of the findings are presented in two main sections: discussions of the descriptive phase of the study, and discussions of the evaluation phase of the study as follow.

7.3.1 Discussions of descriptive phase of the study

7.3.1.1 The utility of frequent words and word lists in language teaching

Developing a medical academic word list of textbooks and considering their differences in occurrences across subdisciplines of medicine, i.e. internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery, besides finding the most frequent collocation/phraseological patterns of the words included in the MAWLOT are the emphases of the first three research questions as the descriptive phase of the study. Some researchers believe that the results of developing specialised word lists show that vocabulary use often varies significantly across different academic disciplines (e.g. Lei & Liu, 2016) hence "it is important to develop discipline-specific academic vocabulary

lists" (p. 42). Regarding to this issue, there are some academic word lists which have been developed to serve the students' needs in academic areas (see section 2.4). Among the criteria that should be revealed to consider in developing word lists, the word frequency has been one of the important ones (see section 2.5.1). The utility of frequent words and compilation of specialised word lists in language teaching has been considered by many researchers. For example, according to Aitchison (1987) word frequency is known to be a crucial variable in text comprehension. In addition, frequent words are recognized faster than infrequent words (Jurafsky, 1996, cited in Baker, 2012). They are also more used by learners than infrequent words (Tono, 2002).

From the first word lists compiled for academic purposes (e.g. the American University List (Praninskas, 1972) to the recent ones (e.g. Academic Vocabulary List (AVL) (Gardner & Davies, 2013); the frequency of words has been considered as a crucial variable to develop word lists for academic purposes. Some researchers even define academic vocabulary based on their frequency. For example, academic vocabulary is defined as the vocabulary that is more frequent in academic texts than in general texts (Hagen et al., 2016). Also, academic core words are "those that appear in the vast majority of the various academic disciplines" (Gardner & Davies, 2013, p. 8).

In addition, the importance of applying word lists based on word frequency in learning and teaching language has been mentioned by many researchers. For example, Chanasattru and Tangkiengsirisin (2016) state that although many educational institutions in Thailand introduce new vocabulary to their students to apply for reading and writing in their specific fields, teachers cannot decide which vocabulary is useful to prepare students to read and write research articles. They add that one problem is that "teachers do not know which words frequently appear and are truly representative of the vocabulary in specific fields" (p. 45). Furthermore, studies on the effect of word

frequency show that lists of high-frequency words are better recalled than lists of low-frequency words (Miller & Roodenrys, 2012). It has also been stated that frequency lists are the obvious starting point for preparing a wordlist for deliberate learning (Kerr, 2015). A word list by frequency also "provides a rationale basis for making sure that learners get the best return for their vocabulary learning effort" (Nation, 1997, p. 17). Finally, “word lists lie at the heart of good vocabulary course design, the development of graded materials for extensive listening and extensive reading, research on vocabulary load, and vocabulary test development” (Nation, 2016, p. xi).

As for the utility of word lists in language teaching, it has been stated that “word lists can be a useful tool to help focus our vocabulary teaching” (Lessard-Clouston, 2016, p. 6). Considering existing word lists, though, it reveals that for most learners in specific fields the ideal list does not exist (Kerr, 2015). Either it will have to be created, which requires a significant amount of time and expertise (Timmis, 2015), or available best-fit will have to suffice, many research (including the present study) have been done to develop frequency wordlists to apply in language teaching (see section 2.4). The findings of the present study show that the study has created a university word list of the most frequent words that medical students are exposed to and meet their specific vocabulary needs. Furthermore, to observe the words’ behaviours in the context, the present study has considered the behaviours of the words included in the MAWLOT by studying their collocation/phraseological patterns. The efficacy of using collocation/phraseological patterns in language teaching is discussed in the following section.

7.3.1.2 The utility of collocation/phraseological patterns in language teaching

According to Wu et al. (2016), electronic corpora and corpus-based tools have created new potential for learners to explore multiword units, such as collocations. Smith in 2005

stated that collocation must be included in the curriculum. Durrant (2008) counts collocation as the most significant component of turning passive words into active ones. It has also been suggested by some researchers (e.g. Xue & Nation, 1984; Hoshino, 2010) that teachers have to help students understand collocations and how they relate to particular items from word lists that they use (see also section 6.2). Furthermore, Lessard-Clouston (2013) encourages teachers to teach word lists and their collocations, so that students can see the relationships between words. In addition, according to Hunston and Francis (2000), researchers in language teaching "place importance upon lexical phrases because of their frequency and their importance to a 'nativelike' production of the language" (p. 10). They state that lexical phrases are important because "they allow language learners to produce language that is phraseologically similar to that of native speakers" (p. 10). Furthermore, Nattinger and DeCarrico (1989) point out to the lexical phrases as "patterned sequences" (p. 11).

All the above statements, besides noticing the point that collocations are of great importance for second language learners have encouraged the author of the present study to produce a list of the most frequent patterns of the words included in the MAWLOT. The resulted collocation/phraseological list of the present study, like other subject-specific collocations lists, may be useful in helping students come to know and practice target vocabulary particularly relevant for their academic studies (Lessard-Clouston, 2013).

7.4 Conclusions of the descriptive phase

In conclusion, to response the Research Questions 1-3, a modified list (by removing nonacademic words in medicine) formed the Medical Academic Word List of Textbooks (MAWLOT) with 459 words (see Appendix D). Of the 459 words, 368 words are nouns, 83 words are adjectives, and 8 words are verbs. The findings also

reveal that of the 459 word families in the list, 417 word families (91%) in the MAWLOT occur in 4 or more of the subdisciplines under study. The most frequent patterns of the words were identified as well.

It has also to be noticed that, the headwords of the word families are generally the most frequent words among the word families included in the MAWLOT (just 15 headwords did not meet the criteria for inclusion in the MAWLOT). Furthermore, the findings of the study indicate that there is a positive relationship between frequency and range of the words included in the MAWLOT. The positive amount of the *linear correlation coefficient (r) test* between the frequency and range of the words included in the MAWLOT (r_s= *0.05, p= 0.000*) shows that when the frequencies of words are increased, the probabilities of the words occurring in a wider range of the subcorpora are increased as well.

The differences in the frequency distribution of the words of MAWLOT across the subcorpora of the Medical Academic Corpus were also considered. According to the findings, the MAWLOT appears to be slightly advantageous for pediatrics texts (12.51%) with a similar coverage of the subdiscipline of internal medicine (12.14%). As Table 4.1 shows, pediatrics covers the subject areas of Perinatology, Pediatrics and Child Health, and internal medicine covers 17 or more subject areas, including Cardiology and Cardiovascular Medicine, Clinical Neurology, Endocrinology, Diabetes and Metabolism, Critical Care, Intensive Care. The high coverage of the subcorpus of internal medicine by vocabulary included in the corpus can be seen as a strong point for comprehending a mass of different written texts included in the subdiscipline of internal medicine. That is, controlling the words included in the MAWLOT enables medical students fulfil their vocabulary needs at least in 17 subjects of medicine, which is very helpful for students. This statement is also true for the coverage of the other three

subdisciplines of surgery, obstetrics and gynecology, and psychiatry (10.74%, 11.78%, and 9.21%).

It should also be noted that different subdisciplines display different frequencies of the words related to the subject area. For example, the word *maternal* (of or relating to a mother, especially during pregnancy or shortly after childbirth) occurs with a high frequency in both pediatrics and obstetrics and gynecology, while the word *mania* (mental illness marked by periods of great excitement, euphoria, delusions, and overactivity) occurs more often in the subdiscipline of psychiatry. The word *mesenteric* (the double layer of peritoneum suspending the intestine from the posterior abdominal wall) also occurs more often in the subdiscipline of surgery than in the other subcorpora.

In addition, the present study found that the role of applying medical technical words to fulfil the medical learners' vocabulary needs in reading medical textbooks is considerable. As noted in the literature review, technical words with about 5% of the running words in a text besides the high-frequency (the GSL) (80%) and academic words (10%) can enable student to get closer to the 95% threshold of text coverage, as Coxhead and Nation (2001) mention. This statement also applies to the present study by presenting a medical word list of textbooks besides a list of medical technical words (53 words) (see Table 5.3). Finally, the findings of the descriptive phase support the idea that developing an academic word list pertaining to different fields and subject areas is a very useful (and possibly necessary) way to fulfil the learners' vocabulary needs of academic texts in a given field (Lei & Liu, 2016). Discussions on the utility of teaching frequent collocation/phraseological patterns in L2 have been presented in Chapter 6.

7.5 Discussions of evaluation phase of the study

7.5.1 Evaluation of the coverage of the MAWLOT

Measuring the coverage of included words in the corpus is the first step to evaluate any word list, according to Coxhead (2000). Research Questions 4 - 6 consider the evaluation phase of the present study to assess the coverage of the Medical Academic Corpus beyond the GSL (West, 1953) and the AWL (Coxhead, 2000), the coverage of another medical academic corpus, and the comparison of MAWLOT with the AWL and the MAWL (Wang et al, 2008) in terms of types, tokens, coverage and running words of each word list. This comparison verifies that the MAWLOT developed in the present study is an improvement on existing word lists. In other words, this phase of the study intended to show the representativeness of the MAWLOT in terms of its high coverage in medical academic English texts against the low coverage of the AWL (Coxhead, 2000) and the MAWL (Wang et al, 2008) of the texts of research articles to address which list can better serve the vocabulary needs of medical users.

The results of the Research Question 4 show that the coverage of the Medical Academic Word List of Textbooks (MAWLOT) is 11.27% of the vocabulary of the Medical Academic Corpus, which is good coverage to assist medical students in their vocabulary needs in reading medical texts. The coverage of the Academic Word List (AWL) is 10% of the vocabulary in the Academic Corpus (Coxhead, 2000) (see Table 5.7). However, the coverage of the AWL is good (10%), but considering the subcorpora of the Academic Corpus, it appears that the corpus with about 3.5 million running words does not contain any medical texts, and consequently, the word list retrieved from the corpus does not include any medical words. Therefore, medical students cannot use the AWL as a vocabulary reference for recognising which words are valuable to assist them in fulfiling their vocabulary needs in reading medical texts. The

Academic Corpus used to develop the AWL contained four disciplines of arts, law, science and commerce with coverage of 9.3%, 9.4%, 9.1%, and 12.0%.

In addition, in the study of Wang et al. (2008), which developed the Medical Academic Word List (MAWL) of medical research articles, the coverage of the word list is 12.24% (see Table 5.7) of the vocabulary included in the corpus under their study. Considering details of the corpus in the study of Wang et al. (2008), it appears that the corpus is a written specialized corpus containing 1,093,011 running words from 288 written texts of medical research articles. The texts included in the corpus were selected from 1800 journals in 24 disciplines from natural science to social science, medicine and dentistry in 32 subject areas (Wang et al., 2008). In contrast with Wang et al.'s (2008) study, the Medical Academic Corpus of the present study contains just medical written texts (see Appendix A and Table 4.5) from the medical textbooks to serve the medical students' vocabulary needs in their reading medical texts.

Research Question 5 specifically considered the coverage of MAWLOT by examining the vocabulary of a second corpus and considering differences on the coverage between AWL, MAWL, and MAWLOT. With regard to examining AWL, MAWL, and MAWLOT in another corpus, there is no report of examining the MAWL (Wang et al., 2008) in a second corpus beyond the main corpus of that study. But, in examining the AWL in a second corpus, Coxhead (2000) reports that the resulting coverage of the AWL of vocabulary of the second corpus is 8.5%. Coxhead believes that the low coverage of the AWL of the second corpus in her study could be partly a result of the large proportion of science texts in the corpus. The second corpus in the study of Coxhead was not a balanced corpus since it contained 82,000 running words in arts, 53,000 in commerce, 143,000 in law, and 400,000 in science (Coxhead, 2000), for a total of 678,000 running words. Therefore, to examine the word list of the present study (MAWLOT) in a second corpus, I compiled a corpus with the same criteria and

sources in order to include texts in the main corpus. In comparison with the study of Coxhead (2000), the second corpus of the present study was a balanced corpus with about 70,000 running words of each subcorpus (see Table 4.5), that is, 350,000 tokens in all. The resulting coverage of MAWLOT of the second corpus vocabulary is 10.48%, which is higher than the coverage of the AWL of the vocabulary of the second corpus (8.5%) in Coxhead's (2000) study.

In sum, while the coverage of the AWL of vocabulary of the second corpus is just 8.5%, the good coverage of the MAWLOT of the vocabulary of the second medical academic corpus (10.48%) confirms that in the absence of a special word list based on medical textbooks the MAWLOT can be used by medical users in their reading medical textbooks.

7.5.2 Evaluation of the comparison of MAWLOT with the AWL and the MAWL

Research Question 6 aimed to find out which list (the AWL, the MAWL, and the MAWLOT) is more beneficial to be used by medical learners in their vocabulary needs in reading medical textbooks.

The AWL, one of the most well-known academic word list, has recently been the subject of some discussion (Lei & Liu, 2016) in its usefulness for all academic domains. For example, Hyland and Tse (2007), and Ward (2009) state that teachers and materials developers should not mislead students into believing that there is just a single collection of words (i.e. the AWL) to learn and transfer across different fields of study. But, they should guide students to apply a word list, which contains high frequency words, academic and specialised/technical vocabulary, in order to guide them through their academic studies (Lei & Liu, 2016). In recognition of a need for academic vocabulary lists for various specific disciplines, several discipline-specific academic word lists have been developed (see section 2.4). The present study has also been conducted to respond to the need of students in a specific discipline such as medicine.

Furthermore, when comparing the Medical Academic Word List (MAWL) of Wang et al. (2008) with MAWLOT, a special note should be made regarding the differences between the Medical Academic Corpus of the present study and the corpus used in Wang's et al. study. First, the Medical Academic Corpus is exclusively composed of medical textbooks while the corpus used to develop the MAWL consists of research articles with the IMRD (Introduction - Method - Result - Discussion) structure. The importance of including textbooks in any corpus used to develop a word list for students has been mentioned in some studies (e.g. Lei & Liu, 2016). The rationale for using a corpus of medical textbooks in the present study is that the target users of the study are medical students who generally have to read medical textbooks as the main sources in their study while the users of research articles are mostly professionals (Lei & Liu, 2016). Therefore, it is vital to include medical textbooks to the corpus of the study as a source to ensure that the words selected to include in the MAWLOT are indeed those the medical students will encounter in their study.

Second, the Medical Academic Corpus of the present study is much larger (3.5 million running words) than the Wang's et al. corpus of 1,093,011 running words, which makes the data and the vocabulary selected of the present study much more representative and reliable (Lei & Liu, 2016) than the Wang's et al. study. It is even comparable in size to the corpus that Coxhead (2000) used in developing the general AWL with 3.5 million running words. It is also necessary to note that, according to Lei and Liu (2016), the MAWL contains some GSL words (15 words) (e.g. *Background*, *entry*, *formation*, *independent*, etc.), while the GSL's words have been removed from the MAWLOT to avoid the overlap between them and medical academic words (see Appendix E).

7.6 Conclusions of the evaluation phase

Regarding the coverage of MAWLOT of the vocabulary of the Medical Academic Corpus, the findings suggest that employing the MAWLOT to fulfil medical students' vocabulary needs of medical written texts may be necessary for medical students. The good coverage of MAWLOT in covering medical textbooks other than the texts included in the Medical Academic Corpus is a good testimony to this statement. Furthermore, although the students might not learn all the words and their collocations, it is still beneficial to know the words and their different collocations in medical contexts to raise students' awareness of the importance of word study. This awareness can bring insight to word use, which will in turn motivate students to use the corpus information for their reading medical textbooks.

In addition, based on the results of the comparative analysis, the MAWLOT boasts a much better coverage of medical English textbooks than the existing medical academic word list developed by Wang et al. (2008). Thus, this list should better serve the needs of medical English learners in reading textbooks. This study also provides evidence for the need to include, in discipline-specific vocabulary lists, and collocation/phraseological patterns in medicine. Therefore, the Medical Academic Word List of Textbooks (MAWLOT) provides students with a chance to expand their vocabulary for reading medical texts. Obviously, as already stated in the literature review chapter, in EAP (as in any other type of formal teaching) incidental vocabulary learning is very ineffective and students are not likely to pick up the academic vocabulary automatically or effortlessly (Martin, 2009). Studying the words included in the MAWLOT and their collocations to increase the vocabulary knowledge in reading written medical texts is more effective than studying medical texts incidentally in enhancing students' word knowledge. The findings can also provide useful pedagogical

implications for language learning by employing an academic word list in EAP situations, especially with non-native medical students.

7.7 Pedagogical implications

Providing the best possible instruction for students is a desire of many instructors, which is achievable by using corpus linguistics in classrooms (Bennett, 2010). One of the major results of corpus-based studies is developing wordlists in a specific discipline to help students practice the frequently used words of the discipline in everyday classrooms (Coxhead, 2010). One way to approach the question of what vocabulary to teach is to consider existing principled lists of English words, research related to their use, and students' interests, needs, and goals (Lessard- Clouston, 2013).

The Medical Academic Word List of Textbooks (MAWLOT) is the result of a corpus-based lexical study that "create lists, concordances or data regarding the clustering of linguistic items in coherent, purposeful texts" (Coxhead, 2000, p. 227). The vocabulary lists, developed from the corpus-based studies are used to teach the different areas of vocabulary, like teaching the academic vocabulary (e.g. Coxhead, 2000) and collocations (e.g. Ellis, Simpson-Valch, & Maynard, 2008). Using corpora to develop wordlists based on frequency is mostly because the Nonnative English Speakers' (NNES) need to learn specialised vocabulary for different disciplines (Thompson, 2006). The word lists contain essential words (Fisher & Frey, 2008) for a specific discipline that present information on word sequences, frequency and range of single words analyzed by computer tools from a large corpus (Cobb & Horst, 2004). In academic areas, the Academic Word List (AWL) developed by Coxhead (2000) has been described as extremely useful for identifying words likely to help students to comprehend academic texts (Nagy & Townsend, 2012).

The findings yielded by the present study also imply that identifying academic vocabulary for instruction, i.e. specifying words that are valuable to teach and learn, is

essential. With respect to the importance of selecting academic vocabulary for instruction, Folse (2004) noted that teachers should choose a list related to their English language learners' (ELLs') objectives (such as conversation, academic work, or business) and proficiency level. Graves (2006, 2008) also proposed that a teacher should compare the words in the text students are reading with the words on a word list, such as: (a) *A New Academic Word List (AWL) (Coxhead, 2000)*; (b) *Building Academic Vocabulary: Teacher's Manual (Marzano & Pickering, 2005)*; (c) *The First 4,000 Words (for basic words) (Graves., 2008)*; (d) *Words Worth Teaching: Closing the Vocabulary Gap (Biemiller, 2010) (for less basic words)*, etc. In medicine, the MAWLOT developed in the present study can be used as a reference wordlist in EAP/EMP courses to select the appropriate words for teaching and learning textbooks. Using a reference wordlist such as MAWLOT allows vocabulary learners to learn the words explicitly or intentionally by direct access to the words and their collocations.

In addition, the MAWLOT could be an adequate resource for learners and teachers of medical English as well as for syllabus designers in medical English education. Lewis (2000) points to the role of teachers as a guide to help students to choose the right vocabulary and word list in their learning. As has been explained in the study (see section 1.3), MAWLOT contains only the most frequent medical academic words found in textbooks. Therefore, the list can help medical learners, teachers, and syllabus designers to focus their attention on only the most useful and frequent medical vocabulary items to enable learners to master the vocabulary they need to understand medical texts they read. After the learners have understood the MAWLOT's words, they can move onto the medical technical list of the study to understand more specialised medical texts. The list is unique, in which none of the studies in developing a medical academic word list has provided such a technical list. Medical technical words have

been even "excluded" from the study of Lei and Liu (2016, p. 55) in developing medical academic vocabulary list (MAVL) based on both research articles and textbooks.

Furthermore, teachers should encourage students to pay attention to the collocation/phraseological patterns (Chung & Nation, 2004; Hyland & Tse, 2007) to understand and learn the use of academic words (Lei & Liu, 2016). By providing a list of the most frequent collocations of the MAWLOT's words, this study follows this belief that it is better to present new words as collocations (Lewis, 2000) at least with one or two more items (Reza & Ashouri, 2016). Therefore, it is recommended that teachers familiarise their students with the notion of collocation by teaching different types of collocation through different exercises to improve EAP/EMP learners' knowledge of vocabulary.

An academic word list such as the MAWLOT is compiled in a principled way and should be learned in a principled way (Coxhead, 2010). Often times, however, while learning can start with the direct study of the words included in the word list, learners might never practice the words in adequate ways. For example, they may just focus on the spelling or meaning of words included in the list. In addition, highlighting the MAWLOT's words in dictionaries, especially in university-level dictionaries, could help students interested in medical academic vocabulary to find the words more quickly and easily. Including MAWLOT in glossaries of medical textbooks or in academic websites like the Compleat Lexical Tutor (Cobb, 2011) also could help learners to discover which words from the MAWLOT are usually found in pertinent texts. MAWLOT could also be used in testing. Similar to the Vocabulary Levels Test, which contains the AWL's words in a section to assess the students' ability to comprehend academic texts (Coxhead, 2010), instructors may include a section that uses the

MAWLOT for assessment of first-year medical students' ability to comprehend the texts used in their courses.

7.8 Recommendations for further study

Developing and using corpora to obtain information about vocabulary frequency in registers of interests in language teaching hold promise in corpus-based studies (Coxhead, 2000). Future research might effectively build on the findings of the current corpus-based lexical study in the following ways:

1. Compare the obtained findings from the Medical Academic Corpus with those from the larger corpora such as those used in medical dictionary-making. Having a balanced corpus with an equal number of running words per discipline would help to have a full comparison. Using the last references to develop the medical academic corpus might also be helpful in obtaining new information.

2. Investigate to what extent the medical technical list (53 words) of the present study will serve the medical students in comprehending texts.

3. Consider, in more depth, what are the grammatical behaviours of the most frequent collocations of words included in the MAWLOT.

4. Obtain more in-depth information about medical academic vocabulary. Do the words in the MAWLOT have roughly the same meaning in different disciplines of medicine? Do some of the lexical items take on a grammatical function in medical texts?

5. Recheck the MAWLOT in larger corpora or in other genres of medicine, such as spoken medical academic English, to investigate if the MAWLOT, which is based on the written medical academic English, accounts for spoken medical academic English, or spoken medical academic English needs a separate medical academic word list.

6. Compare collocation/phraseological patterns of the most frequent words of the MAWLOT with the collocation/phraseological patterns of the least frequent words revealed by the study.

7. It would also be useful for future research to develop various specialist area specific medical vocabulary lists to meet medical specialists' needs in L2 medical learners.

In summary, this study presented a Medical Academic Word List of Textbooks, i.e. the MAWLOT, a frequent collocation list of the words included in the MAWLOT, and the differences in the coverage of the MAWLOT, the AWL, and the MAWL to present evidences in the usefulness of MAWLOT, as a proper word list to be applied by medical learners in reading medical textbooks to fulfil their vocabulary needs. This study also encourages medical teachers to use an appropriate word list (i.e. MAWLOT) for vocabulary teaching in medical context. Understanding which list is more adequate in reading medical textbooks is useful to help fill in students' gaps in their vocabulary knowledge (Lessard- Clouston, 2013).

References

Ackerman, K. & Chen, Y-H. (2013). Developing the Academic Collocation List (ACL): A corpus- driven and expert-judged approach. *Journal of English for Academic Purposes,* 12, 235-247.

Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.

Aluthman, E. S. (2017). Compiling an OPEC word list: A corpus-informed lexical analysis. *International Journal of Applied Linguistics & English Literature,* 6 (2), 78-92.

Allwright, R. L. (1984). The Importance of Interaction in Classroom Language Learning. *Applied linguistics*, 5 (2), 156-171.

Altman, D. (1991). *Practical statistics for medical research.* London: Chapman Hall.

Anthony, L. (2012). AntConc (version 3.3.5). Tokyo: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp/

Aston, G. (2002). The learner as corpus designer. *Language and Computers*, 42 (1), 9-25.

Atkins, S. & Clear, J. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7 (l), 1-16.

Baker, p. (2012). *Contemporary corpus linguistics*. London/New York: Continuum.

Bauer, L., & Nation, I. S. P. (1993). Word Families. *International Journal of Lexicography.* 6 (4), 253-279.

Bennett, G. (2010). *Using Corpora in the Language Learning Classroom*. Michigan: Michigan University Press.

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8 (4), 243-257.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14 (3), 275-311.

Biemiller, A. (2010). *Words worth teaching: Closing the vocabulary gap*. Columbus: McGraw-Hill SRA.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. New York: Routledge.

Brown, R. (1973). *A first language: The early stages*. Boston: Harvard University Press.

Brooker, C. (2008). *Churchill Livingstone medical dictionary*. Amsterdam: Elsevier

Health Sciences.

Campion, M. E., & Elley, W. B. (1971). *An academic vocabulary list*. Wellington: New Zealand Council for Educational Research.

Candlin, C. N. (1981). *The communicative teaching of English: Principles and an exercise typology*. Harlow: Longman publishing group.

Chanasattru, S. & Tangkiengsirisin, S. (2016). Developing of a High Frequency Word List in Social Sciences. *Journal of English Studies,* 11, 41-87.

Chandler-Burns, R. M. (1997). Non-native writing instruction for medical purposes. In *Meeting of teachers of English to speakers of other language.*Aizuwakamatsu: University of Aizu.

Chen, Q., & Ge, G. C. (2007).A Corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes*, 26, 502-514.

Cheng, W. (2011). *Exploring corpus linguistics: Language in action*. New York: Routledge.

Chomsky, N. (1975). *Formal properties of grammars.* New York: Wiley.

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a foreign language*, 15 (2), 103-116.

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12 (4), 345-360.

Cobb, T., Windows, P., & Free, P. (2005). The Compleat Lexical Tutor (Version 4). Retrieved from http://tesl-ej.org/ej31/m2.html

Cobb, T. (2011). The Compleat Lexical Tutor (Version 6.2). Retrieved from http://www.lextutor.ca/

Cobb, T. (2014). The Compleat Lexical Tutor (Version 8). Retrieved from http://www.lextutor.ca/

Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In B. Laufer & P. Bogaards (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15–38). Amsterdam: John Benjamins.

Cohen, A., Glasman. H., Rosenbaum-Cohe, P.R., Ferrara, J. & Fine, J. (1988). Reading English for specialized purposes: Discourse analysis and the use of standard informants. In P, Carrell, J. Davine. & D. Skey (Eds.), *Interactive approaches to second language reading (*pp. 152-167). Cambridge: Cambridge University Press.

Cobuild, C. (1995). *English language dictionary.* London: Collins.

Cortés, K. H. (2002). Youth and the study of foreign language: An investigation of attitudes. *Foreign Language Annals*, 35 (3), 320-332.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly,* 32 (2), 213-238.

Coxhead, A. (2010). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, 45 (2), 355-362.

Coxhead, A., & Nation. I. S. P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.). *Research perspectives on English for academic purposes* (pp. 252-267). Cambridge: Cambridge University Press.

Douglas, B. H. (2000). *Principles of language learning and teaching*. Harlow: Longman.

Dudley-Evans, T., & St. John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge university press.

Durrant, P. (2008). *High frequency collocations and second language learning*. Ph.D. dissertation, University of Nottingham: UK.

Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., et al. (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *PLOS One*, *11* (3).

Ellis, N. C. (1998). Emergentism, connectionism, and language learning. *Language Learning*, 48, 631– 664.

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in second language acquisition*, 27 (2), 305-352.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42 (3), 375-396.

Farjami, H. (2014). Key lexical chunks in applied linguistics article abstracts. *Journal of Teaching Language Skills*, 6 (3), 51-73.

Farjami, H. (2016). Good reasons to use psychological questionnaires in enhancing teaching English as a foreign language. *Journal of Creative Practices in Language Learning and Teaching (CPLT),* 4 (1), 40-49.

Farrell, P. (1990). *Vocabulary in ESP: A lexical analysis of the English of electronics and a study of semi-technical vocabulary.* CLCSO occasional paper No. 25, Dublin: Trinity College.

Firth, J. R. (1957). *A synopsis of linguistic theory*: *1930-1955*. Oxford: The philological society.

Fisher, D., & Frey, N. (2008). Releasing responsibility. *Educational Leadership,* 66 (3), 32-37.

Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom* teaching. Michigan: University of Michigan Press.

Fox, G. (1998). The Internet: Making it work in the ESL classroom. *The Internet TESL Journal,* 4 (9), 1-4.

Francis, W. N., & Kucera, H. (1979). *Brown corpus manual*. Rhode Island: Brown University.

Francis, G., Hunston, S., & Manning, E. (1996). *Collins COBUILD grammar patterns 1: verbs*. London: Harper Collins Publishers.

Francis, G., Hunston, S., & Manning, E. (1998). *Collins COBUILD grammar patterns 2: Nouns and adjectives*. London: Harper Collins Publishers.

Frank, R. A. (2000). Medical communication: non-native English speaking patients and native English speaking professionals. *English for Specific Purposes*, 19 (1), 31-62.

Fraser, S. (2007). Providing ESP learners with the vocabulary they need: Corpora and the creation of specialized word lists. *Hiroshima Studies in Language and Language Education*, 10, 127-143.

Fraser, S. (2009). Breaking down the divisions between general, academic and technical vocabulary: The establishment of a single, discipline-based word list for ESP learners. *Hiroshima Studies in Language and Language Education,* 12, 151-176.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35* (3), 305-327.

Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles, *Teaching and language corpora* (pp. 83-99). Harlow: Longman.

Ghadessy, M. (1979). Frequency counts, word lists, and materials preparation: A new approach. *English Teaching Forum*, 17 (1), 24-27.

Ghezelseflou, M. & Seyedrezaei, S. H. (2015). The effect of teaching vocabulary through collocations on the vocabulary learning of Iranian EFL students. *Journal of Language Sciences and Linguistics*. 3 (1), 9-16.

Gläser, R. (1988). The grading of idiomaticity as a presupposition for a taxonomy of idioms. In W. Hüllen and R. Schulze (Eds.), *Understanding the Lexicon* (pp. 264-279). Tübingen, Max Niemeyer.

Gledhill, C. J. (2000). *Collocations in science writing* (V. 22). Scotland: Gunter Narr Verlag.

Godwin-Jones, B. (2001). Emerging technologies: Language testing tools and technologies. *Language, Learning & Technology*, 5 (2), 8-12.

Graham, J. G., & Beardsley, R. S. (1986). English for specific purposes: Content, language, and communication in a pharmacy course model. *TESOL Quarterly*, 20 (2), 227-245.

Graves, M. (2006). *The vocabulary book: learning and instruction.* New York: Teachers College press.

Graves, M. (2008). The language curriculum: A social contextual perspective. *Language Teaching*, 41 (2), 147-181.

Hagen, K. Johannessen, J. B. & Saidi, A. (2016). Constructing a Norwegian Academic Wordlist. *Tenth international conference on language resources and evaluation of LREC Proceedings*, 1457-1462.

Hajiyeva, K. (20016). A corpus-based lexical analysis of subject-specific university textbooks for English majors, *Ampersand*, *2*, 136-148.

Halliday, M. A. (1993). Towards a language-based theory of learning. *Linguistics and education*, 5 (2), 93-116.

Heatley, A., & Nation, P. (1996). *Range [Computer software]*. Wellington: Victoria University of Wellington.

Hoshino, Y. (2010). The categorical facilitation effects on L2 vocabulary learning in a classroom setting. *RELC Journal*, *41* (3), 301-312.

Howarth, P. (1998). The phraseology of learners 'academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp.161-186). Oxford: Clarendon Press.

Hsu, J. Y. (2007). Lexical Collocations and Their Impact on the Online Writing of Taiwanese College English Majors and Non-English Majors. *Online Submission*.

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*, 33, 54-65.

Hsu, J. Y & Hsu, L. C. (2007). Teaching lexical collocations to enhance listening comprehension of English majors in a technological university of Taiwan. *Soochow Journal of Foreign Languages & Cultures*, 24, 1-32.

Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In A.E. Farstrup & S.J. Samuals (Eds.). *What research has to say about vocabulary instruction*, (pp.106-129). Newark: International Reading Association.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge

University Press.

Hunston, S., & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins: Amsterdam/Philadelphia.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. London: Cambridge University Press.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Michigan: University of Michigan Press.

Hyland, K., & Hamp-Lyons, L. (2002). EAP: Issues and directions. *Journal of English for academic purposes*, 1 (1), 1-12.

Hyland, K., & Tse, P. (2007). Is there an academic vocabulary? *TESOL Quarterly*, 41, 235-253.

Johansson, S. (1980). The LOB corpus of British English texts: Presentation and comments. *ALLC Journal*, 1 (1), 25-36.

Johns, A. M. (1990). L1 composition theories: Implications for developing theories of L2 composition. In *Second language writing: Research insights for the classroom* (pp. 24-36).

Jones, C. & Waller, D. (2015). *Corpus linguistics for grammar: A guide for research*. London/New York: Routledge.

Juffs, A. (2009). Second Language Acquisition of the Lexicon. In Ritchie, William & Bhatia, Tej K. (Eds.), *The New Handbook of Second Language Acquisition*, Bingley: Emerald Group Publishing Limited.

Kerr, P. (2015). Adaptive learning. *ELT Journal*, 70 (1), 88-93.

Kennedy, G. (2014). *An introduction to corpus linguistics*. New York: Routledge.

Kilgarriff, A., & Grefenstette, G. (2003).Introduction to the special issue on the web as corpus. *Computational linguistics*, 29 (3), 333-347.

Kjellmer, G. (1984). Some thoughts on collocational distinctiveness. In J. Aarts &W. Meijs (Eds.), *Corpus Linguistics* (pp.163-171). Amsterdam: Rodopi.

Kuo, C. (2009). An analysis of the use of collocation by intermediate EFL college students in Taiwan. *Arecls*, 6, 141-155.

Lam, J. K. M. (2001). *A study of semi-technical vocabulary in computer science texts, with special reference to ESP teaching and lexicography* (Research reports, Vol.3). Hong Kong: Hong Kong University of Science and Technology, Language Centre.

Laerd Statistics, (2017). *Linear Regression Analysis using SPSS Statistics*, https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, 16 (3), 307-322.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language testing*, 16 (1), 33-51.

Leech, G. (1992). *Corpora and theories of linguistic performance: Directions in corpus linguistics.* Berlin: Mouton de Gruytor.

Leffa, V. J. (1998). Textual constraints in L2 lexical disambiguation. *System, 26* (2), 183-194.

Lessard-Clouston, M. (2013). Word lists for vocabulary learning and teaching. *CATESOL Journal*, 24 (1), 287-305.

Lessard-Clouston, M. (2016, November 18). *New Word Lists for Vocabulary Teaching.* Paper presented at 47th Annual CATESOL Conference, San Diego: U.S.A.

Lehecka, T. (2015). Collocation and colligation. In *Handbook of Pragmatics Online*. Amsterdam: John Benjamins Publishing.

Leeke, P., & Shaw, P. (2000). Learners' independent records of vocabulary. *System*, 28 (2), 271-289.

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53.

Lewis, M. (2000). Materials and resources for teaching collocation. In *Teaching collocation: Further development in the lexical approach* (pp.186-204). Hove: Language Teaching Publications.

Li, E. S. L., & Pemberton, R. (1994). An investigation of students' knowledge of academic and subtechnical vocabulary. In *Proceedings joint seminar on corpus linguistics and lexicology, Guangzhou and Hong Kong* (pp. 183-196). Hong Kong: HKUST Language Center.

Li, Y., & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, 38 (3), 402-411.

Liu, C. P. (2000). A study of strategy use in producing lexical collocations. *Selected papers from the tenth International Symposium on English Teaching*, 481-492. Taipei: Crane.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. *Handbook of second language acquisition*, 2, 413-468.

Lorge, I. (1944). Predicting readability. *The Teachers College Record*, 45 (6), 404-419.

Lynn, R.W. (1973). Preparing word lists: a suggested method. *RELC Journal*, 4 (1), 25-32.

Mahlberg, M. (2005). *English general nouns: A corpus theoretical approach*. Amsterdam: John Benjamins Publishing.

Mahlberg, M. (2006). Lexical cohesion: Corpus linguistic theory and its application in English language teaching. *International Journal of Corpus Linguistics*, 11 (3), 363-383.

Martin, J. R. (2009). Genre and language learning: A social semiotic perspective. *Linguistics and Education*, 20 (1), 10-21.

Martyńska, M. (2004). Do English language learners know collocations? *Investigationes linguisticae*, 11, 1-12.

Marzano, R. J., & Pickering, D. J. (2005). *Building academic vocabulary: Teacher's manual*. Alexandria: Association for supervision and curriculum development.

McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.

McEnery, T., & Gabrielatos, C. (2006). *English Corpus Linguistics: The handbook of English linguistics*. Oxford: Blackwell Publishing.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York: Routledge.

McKenny, J. A. (2006). *A corpus-based investigation of the phraseology in various genres of written English with applications to the teaching of English for academic purposes* (Doctoral dissertation). Retrieved from http://etheses.whiterose.ac.uk/346/

McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, 49 (1), 25- 42.

Mehring, J.G. (2005). *Developing vocabulary in second language acquisition: From theories to the classroom*. Retrieved on April 15, 2014 from http://www.hpu.edu/CHSS/LangLing/TESOL/ProfessionalDevelopment/200680TWPfall06/03Mehring.pdf.

Mellow, D., & Stanley, K. (2002). Theory development in applied linguistics: Toward a connectionist framework for understanding second language acquisition. *Issues in applied linguistics*, 13 (1), 3- 40.

Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

Miller, L. M. & Roodenrys, S. J. (2012). Serial recall, word frequency, and mixed lists: The influence of item arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38 (6), 1731-1740.

Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol.45). London: Short Run Press Ltd.

Mukaka, M.M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Melawi Medical Journal*, 24 (3), 69–71.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25 (2), 235-256.

Murison-Bowie, S. (1996). Linguistic corpora and language teaching. *Annual review of applied linguistics*, 16, 182-199.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24 (3), 262-282.

Nagy, W. & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47 (1), 91-108.

Nakata, T. A. T. S. U. Y. A. (2006). Implementing optimal spaced learning for English vocabulary learning: Towards improvement of the low-first method derived from the reactivation theory. *The JALT Call Journal*, 2 (2), 3-1.

Nation, I.S.P. (1990). *Teaching and Learning Vocabulary.* New York: Newbury House.

Nation, I.S.P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, 9 (2), 6-10.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamines.

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, 14, 6-19.

Nattinger, J. R., & DeCarrico, J. S. (1989). Lexical phrases, speech acts and teaching conversation. *Aila Review*, 6, 118-139.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied linguistics*, 24 (2), 223-242.

Noguchi, J. 2004. A genre analysis and mini-corpora approach to support professional writing by nonnative English speakers. *English corpus studies*, 11, 101-110.

O'Dell, F. (1997). Incorporating vocabulary into the syllabus. In Schmitt & McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp.258-278). Cambridge: Cambridge University Press.

Palmer, H.E. (1933). *Second interim report on English collocations*. Tokyo: Institute for Research in English Teaching.

Partington, A. (1998). *Patterns and meanings: Using corpora for English Language research and teaching.* Amsterdam: John Benjamin.

Praninskas, J. (1972). *American university word list*. London: Longman.

Rebek-Nagy, G., Warta, V., & Barron, J. P. (2008).Development of international standards for medical communications in English. *Chest*, 134 (3), 661-662.

Reza, Kh, & Ashouri, Sh. (2016). The effect of lexical collocation instruction on intermediate EFL learners' vocabulary size. *Indian journal of fundamental and applied life sciences*. 6 (1), 416-425.

Richards, J. C., & Rodgers, T. S. (2001). Approach and methods in language teaching: A description and analysis. *Cambridge: Cambridge University*.

Sadeghi, K. (2009). Collocational differences between L1 and L2: Implications for EFL learners and teachers. *TESL Canada Journal*, *26* (2), 100-124.

Salager-Meyer, F. (1990). Metaphors in medical English prose: A comparative study with French and Spanish. *English for Specific Purposes*, 9 (2), 145-159.

Sarvari, S., Gukani, A. H., & Khomami, H. Y. (2016). Teaching collocations: Further developments in L2 speaking fluency. *Journal of Applied Linguistics and Language Research*, *3* (3), 278-289.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamin Press.

Shabani, G. (2016). A comparative study on the effect of contrived vs. authentic texts on Iranian intermediate EFL learners' lexical collocations learning. *Journal of Studies in Education*, 6 (1), 20-36.

Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT journal*, 62 (4), 339-348.

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

Sinclair, J. (1996). *Corpus to corpus: A study of translation equivalence*. Oxford: Oxford University Press.

Sinclair, J. (2001). *Small corpus studies and ELT: Theory and practice*. Amsterdam: John Benjamines.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. New York: Routledge.

Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and language*, 83 (2), 227-236.

Smith, C. (2005). *The Lexical Approach: Collocation in High School English*

Language Learners. Oregon: George Fox University.

Stahl, S. A. (1986). Three principles of effective vocabulary instruction. *Journal of Reading*, 29 (7), 662-668.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of educational research*, 56 (1), 72-110.

Stedman, T. L. (2012). *Stedman's Medical Dictionary for the health professions and nursing*. Philadelphia: Lippincott Williams & Wilkins.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taiwo, R. (2004). Helping ESL learners to minimize collocational errors. *The Internet TESL Journal*, 10 (4).

Taylor, C. (2008). What is corpus linguistics? What the data says. *ICAME Journal*, 32, 179-200.

Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10 (1), 1-13.

Teubert, W., & Krishnamurthy, R. (2007*). Corpus linguistics: Critical concepts in linguistics*. New York: Routledge.

Thabane, L. Ma, J., Chu, R., Cheng, J. Ismaila, Lorena, A., Rios, L. P., et al. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical research methodology,* 10 (1), 1-11.

Thabane, L., Hopewell, S., Lancaster, G. A., Bond, C. M., Coleman, C. L., Campbell, M. J., et al. (2016). Methods and processes for development of a CONSORT extension for reporting pilot randomized controlled trials. *Pilot and feasibility studies, 2* (1), 25.

Thompson, P. (2006). Assessing the contribution of corpora to EAP practice. In Z. Kantaridou, I. Papadopoulou & I. Mahili (Eds.), *Motivation in learning language for specific and academic purposes* (pp.1-19). Macedonia: University of Macedonia.

Thorndike, E. L., & Lorge, I. (1944). T*he teacher's handbook of 30,000 words.* New York: Columbia University Teachers' College Press.

Thurston, J., & Candlin, C. N. (1998).Concordancing and the teaching of the vocabulary of academic English. *English for specific purposes*, 17 (3), 267-280.

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. London/New York: Routledge.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). Amsterdam: John Benjamines.

Tono, Y. (2002). *The role of learner corpora in SLA research and foreign language teaching: The multiple comparison approach* (Doctoral dissertation, University of Lancaster).

Vongpumivitch, V., Huang, J. Y., & Chang, Y. 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*, 28 (1), 33-41.

Walker, C. P. (2011). A corpus-based study of the linguistic features and processes which influence the way collocations are formed: Some implications for the learning of collocations. *TESOL Quarterly*, 45 (2), 291-312.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes*, 27, 442-458.

Ward, A. E. (2009). *A Formative Study Investigating Interactive Reading and Activities to Develop Kindergartners' Science Vocabulary*. California: ProQuest LLC.

Wary, A. (2002). *Formulaic Language and the Lexicon*. New York: Oxford University Press.

Watson Todd, R. (2003). EAP or TEAP? *Journal of English for Academic Purposes*, 2 (2), 147-156.

West, M. (1953). *A general service list of English words.* London: Longman, Green & Co.

Woolard, G. (2000). Collocation: Encouraging learner independence. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 28-46). Hove: Language Teaching Publications.

Wu, W. S. (1996). *Lexical collocations: One way to make passive vocabulary active*. Proceedings of the11th Conference on English Teaching and Learning. Chung Hua Polytechnic Institute: China. doi: 10.1.1.546.2776&rep=rep1&type=pdf.

Wu, S., Li, L., Witten, I., & Yu, A. (2016). Constructing a Collocation Learning System from the Wikipedia Corpus. *Constructing a Collocation Learning System from the Wikipedia Corpus*, 6 (3), 18-35.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language learning and communication*, 3 (2), 215-229.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, 24 (3), 4-12.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring reading through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, 57, 541-572.