



A CORPUS-BASED LEXICAL STUDY OF THE  
FREQUENCY, COVERAGE AND DISTRIBUTION OF  
ACADEMIC VOCABULARY IN ISLAMIC  
ACADEMIC RESEARCH ARTICLES

BY

MAYINUER ABUDUKEREMU

A dissertation submitted in partial fulfilment of the  
requirements for the degree of Master of Human  
Sciences (Teaching English as a Second Language)

Kulliyyah of Islamic Revealed Knowledge  
and Human Sciences

International Islamic University  
Malaysia

JULY 2010

## ABSTRACT

One of the main problems faced by both teachers and learners in English for Academic Purposes pertains to the question of which words are worth focusing during teaching and learning. This corpus-based lexical study aimed to explore the use of words in Coxhead's (2000) 'A New Academic Word List (AWL)' in academic journal articles in the field of Islamic studies. Around 472,621 word corpus, called the Islamic Academic Research Articles (IARA) corpus, was created for this study. The corpus consisted of 66 research articles written in English that were published in more than 10 different Islamic academic journals. In order to simulate rich natural contexts, these articles were selected for being both authentic and academic covering a wide range of topics in the field of Islamic studies. The study found that the most frequent 317 AWL words which occurred in the IARA corpus was only 56% of Coxhead's AWL of 570. This finding points to the need for a special AWL for students of Islamic studies. The findings suggest the need to produce field-specific academic word lists incorporating all frequent academic lexical items necessary for the expression of the rhetoric of the specific research area. The findings also revealed that some of the words which were found in the present study were not found in Coxhead's influential Academic Word List. This may be due to the general nature of the AWL. This suggests that vocabulary needs of students in the Islamic studies are characteristically different from those of students in other disciplines. Researchers and teachers therefore need to deal with these learners separately.

**Key words:** English for Academic Purposes (EAP), Corpus Based Study, Islamic Academic Research Articles (IARA) Corpus, Word Family.

## ملخص البحث

إن إحدى المشاكل الأساسية التي يواجهها الأساتذة والطلبة في اللغة الإنجليزية لأغراض أكاديمية تتعلق بقضية التركيز على الكلمات والمصطلحات التي لها أهمية أثناء التعليم والتعلم. فالجزء الأساسي القائم على دراسة بيانات المفردات يهدف إلى استكشاف واستخدام الكلمات للمؤلف Coxhead (2000م) تحت عنوان "قائمة الكلمات الأكاديمية الجديدة" في مقالات المجلة الأكاديمية فيما يخص الدراسات الإسلامية. وعُينت 472621 كلمة للدراسة، سُميت بمجموعة بيانات "مقالات البحوث الأكاديمية الإسلامية" (IARA)، وتتألف دراسة البيانات من 66 مقالة كتبت باللغة الإنجليزية. نشرت في أكثر من 10 مجلات أكاديمية إسلامية مختلفة. ومن أجل مواكبة السياقات المتجددة في الطبيعة؛ فقد اختيرت هذه المقالات لتكون أصيلة وأكاديمية تشمل نطاق واسع من الموضوعات في مجال الدراسات الإسلامية. وقد كشفت الدراسة أن أغلب الكلمات المكررة 317 "AWL" موجودة في مجموعة (IARA) بنسبة 56٪ من مقالة Coxhead البالغ 570 كلمة. هذه النتيجة تبين مدى الحاجة إلى تخصيص "AWL" لطلاب الدراسات الإسلامية. وتشير النتائج إلى مدى الحاجة لتكوين قسم أكاديمي مخصص لجداول الكلمات التي تتضمن جميع المصطلحات الأكاديمية المتكررة واللازمة للتعبير عن بلاغة مجال البحث المحدد. كما أظهرت النتائج أيضاً أن بعض الكلمات التي وُجدت في هذه الدراسة لم توجد في قائمة الكلمات الأكاديمية لـCoxhead، وربما يرجع ذلك السبب إلى الطبيعة العامة لـ "AWL". وتشير هذه الدراسة إلى أن احتياجات المفردات للطلاب في الدراسات الإسلامية تختلف بشكل يميزهم عن الطلاب في تخصصات أخرى. لذا؛ فإن الباحثين والأساتذة يحتاجون إلى التعامل مع هؤلاء الطلاب بشكل منفصل. الكلمات الأساسية: اللغة الإنجليزية للأغراض الأكاديمية (EAP)، دراسة مستند البيانات، مجموعة بيانات "مقالات البحوث الأكاديمية الإسلامية"، مصدر الكلمة.

## APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Human Sciences (Teaching English as a Second Language)

---

Mohamed Ismail Ahamad Shah  
Supervisor

I certify that I have read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Human Sciences (Teaching English as a Second Language)

---

Haja Mohideen Mohd-Ali  
Examiner

This dissertation was submitted to the Department of English Language and Literature and is accepted as partial fulfilment of the requirements for the degree of Master of Human Sciences (Teaching English as a Second Language)

---

Zahariah Pilus  
Head, Department of English  
Language & Literature

This dissertation was submitted to the Kulliyah of Islamic Revealed Knowledge and Human Sciences and is accepted as partial fulfilment of the requirements for the degree of Master of Human Sciences (Teaching English as a Second Language)

---

Badri Najib Zubir  
Dean, Kulliyah of Islamic Revealed  
Knowledge and Human Sciences

## DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Mayinuer Abudukeremu

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

**DECLARATION OF COPYRIGHT AND AFFIRMATION  
OF FAIR USE OF THE UNPUBLISHED RESEARCH**

Copyright @ 2010 by Mayinuer Abudukeremu All rights reserved.

**A CORPUS-BASED LEXICAL STUDY OF THE FREQUENCY, COVERAGE  
AND DISTRIBUTION OF ACADEMIC VOCABULARY IN ISLAMIC  
ACADEMIC RESEARCH ARTICLES**

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without permission of the copyright holder except as provided below:

1. Any material contained in or derived from this unpublished research may only be used by others in their writings with due acknowledgement.
2. IIUM and its library will have the rights to make and transmit copies (print or electronic) for institutional or academic purposes.
3. The IIUM library will have the right to make and store in retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

Affirmed by Mayinuer Abudukeremu.

.....  
Signature

.....  
Date

## ACKNOWLEDGEMENTS

My deepest gratitude goes to my respected supervisor, Assoc. Prof. Dr. Mohamed Ismail Ahamad Shah for all his support and help. Without his encouragement and suggestions, I would not have been able to write this research. I am indeed grateful to him for his supervision, valuable advice and trust that I can make this academic journey to fruition.

From my inner soul I thank Allah for blessing me with the idea and strength to write. Special thanks to all the lectures who have taught me. The knowledge that you have given me is priceless.

I am greatly indebted to my parents and all my family for all their help. I also acknowledge all the help my friends have given me. Without their generosity, this study would be sorely lacking information. I am indeed deeply appreciative to them, for their love and insistence in speeding up my drive to write this dissertation to completion.

Last but not least, my heartfelt thanks to every person who contributed with either inspirational or actual work with this research and also who had selflessly shared their time, talent and treasure in the interest of such a venture. May Allah always bless them all.

# TABLE OF CONTENTS

Abstract .....	ii
Abstract in Arabic .....	iii
Approval Page .....	iv
Declaration Page .....	v
Copyright Page .....	vi
Acknowledgements .....	vii
List of Tables .....	x
List of Figures .....	xi
List of Abbreviations .....	xii
<b>CHAPTER ONE: INTRODUCTION .....</b>	<b>1</b>
1.1 Background of the Study .....	1
1.2 Research Problem .....	3
1.3 Objectives of the Research .....	6
1.4 Research Questions .....	6
1.5 Significance of the Study .....	7
1.6 Conceptual Definitions of Terms .....	7
1.7 Organization of the Chapters .....	9
<b>CHAPTER TWO: LITERATURE REVIEW .....</b>	<b>10</b>
2.1 Introduction .....	10
2.2 Corpus Linguistics .....	10
2.3 Vocabulary Knowledge and Second Language (L2) Proficiency .....	14
2.3.1 Categories of Vocabulary in Academic Context .....	16
2.4 Some Previous Studies on Academic Words .....	18
2.4.1 Coxhead's Academic Word List .....	19
2.4.2 Previous Studies on Coxhead's AWL in Various Disciplines .....	20
2.5 The Summary .....	23
<b>CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY .....</b>	<b>25</b>
3.1 Introduction .....	25
3.2 Research Design .....	25
3.3 Data Collection .....	26
3.3.1 Choosing and Creating Base Word Lists .....	29
3.4 Sampling Method .....	31
3.4.1 Stratified Random Sampling .....	31
3.5 The Instrument for Analysis .....	33
3.5.1 Data Processing and Analysis .....	33
3.6 Summary .....	34
<b>CHAPTER FOUR: DATA ANALYSIS AND FINDINGS .....</b>	<b>35</b>
4.1 Introduction .....	35
4.1.1 Collected Articles .....	36
4.2 Word Frequencies in the Islamic Academic Articles Corpus .....	36
4.2.1 Token .....	36

4.2.2 Type .....	37
4.3 Comparing Frequencies with Other Studies .....	39
4.3.1 Comparing the Coverage of the AWL with Previous Studies .....	39
4.3.2 Comparing the Combined Coverage of the AWL and the GSL with Previous Studies .....	40
4.4 Frequency of Word Families.....	42
4.5 Comparing Word Families with Other Studies .....	46
4.6 Non-AWL content Word forms in the IARA Corpus .....	58
4.6.1 Using Loan Words in the IARA Corpus .....	60
4.7 Discussion .....	61
4.8 Summary .....	63
<b>CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS .....</b>	<b>65</b>
5.1 Introduction .....	65
5.2 The Study .....	65
5.3 Major Findings .....	66
5.4 Scope and Limitation of the Research .....	68
5.5 Pedagogical Implications of the Study .....	69
5.6 Recommendations .....	71
<b>BIBLIOGRAPHY .....</b>	<b>73</b>
APPENDIX: 1 THE ACADEMIC WORD LIST .....	78
APPENDIX: 2 GENERAL SERVICE LIST .....	83
APPENDIX: 3 VARIOUS KINDS OF WORDS IN ACADEMIC TEXT .....	98
APPENDIX: 4 THE WORDSMITH TOOLS .....	100
APPENDIX: 5 SELECTED ACADEMIC RESEARCH ARTICLES FOR IARA CORPUS .....	103
APPENDIX: 6 SELECTING VOCABULARY: AWL IN IARA CORPUS .....	108
APPENDIX: 7 SOME EXAMPLES OF OTHER WORDS AND LOAN WORDS BEGINNING WITH LETTER “T” THAT OCCURRED IN THE IARA CORPUS .....	123

## LIST OF TABLES

<u>Table No.</u>		<u>Page No.</u>
2.1	Comparison of AWL and GSL Coverage in the Agro Corpus and in Other Corpora	22
3.1	Name of the Journals and Number of Articles in the IARA Corpus	28
3.2	Bauer and Nation's Six-Level Scale	30
4.1	Coverage of Lexical Items in the Islamic Academic Research Article Corpus (Raw Numbers and Percentages)	36
4.2	Frequency of Academic Word Tokens and Types Used in the Islamic Academic Research Article Corpus	39
4.3	Comparison of AWL and GSL Coverage in the Islamic Academic Corpus and in Other Corpora	40
4.4	Comparison of AWL Word Coverage Among the Previous Studies	41
4.5	Frequency and Distribution of Academic Word Families Above the Mean in IARA Corpus	43
4.6	Most Frequent Academic Words Across Disciplines	47
4.7	Comparison of the Co-occurring Words of the Ten Most Frequent Words	48
4.8	Ten Top Words in the Seven Corpora	49
4.9	The 20 Most Frequent Families in Different Corpus	52
4.10	The 20 Most Frequent Families (with Family Members) in Islamic RA Corpus	53
4.11	Comparison Between Three Specific Academic Disciplines	54
4.12	The Top 100 Academic Word Families in the IARA Corpus and in Coxhead's Sublist1	57

## LIST OF FIGURES

<u>Figure No.</u>		<u>Page No.</u>
3.1	Multiple Random Sampling Techniques Applied in this Research	32
4.1	Distribution of Words by Tokens	37
4.2	Distribution of Words by Types	38
4.3	Comparison of AWL and GSL Coverage in the Islamic Academic Corpus and in Other Corpora	41
4.4	Comparison of AWL Word Coverage among the Previous Studies	42
4.5	Distribution of Academic Word Families Above the Mean in the IARA Corpus	43
4.6	Comparing Distributions of Academic Word Families with Other Studies	45

## LIST OF ABBREVIATIONS

IIUM	International Islamic University Malaysia
AWL	Academic Word List
EAP	English for Academic Purposes
ESP	English for Specific Purposes
GSL	General Service List
IMRD	Introduction–Method–Results–Discussion
IARA	Islamic Academic Research Articles
L2	Second language
NNS	Non-native English speakers
SRS	Stratified Random Sampling
TL	Target language
WST	WordSmith Tool

# **CHAPTER ONE**

## **INTRODUCTION**

### **1.1 BACKGROUND OF THE STUDY**

“Lexis is the core or heart of language, but it has long been the Cinderella in language teaching” (Lewis, 1993:89). This observation, however, has been changing constantly. The acquisition of vocabulary has been considered to be a crucial component of learning a language and is the subject of numerous research projects (Coady, Magoto, Hubbard, Graney & Mokhtari, 1993; Laufer & Hulstijn 2001; Nation, 2001, 2006; Laufer, 2003; Read, 2000, 2007), in order to improve communicative potential, fluency and accuracy.

In general, the focus of teaching vocabulary in the primary and secondary schools has been to provide the basic vocabulary that is high-frequency words to students. This is essential at the elementary and intermediate levels of English for general purposes. Yet, at the tertiary level, students should be taught more than those basic words. The rationale for this is that for advanced learners, the ability to express ideas, and write with more confidence and more advanced thought processes in a second language (L2) is greatly facilitated or hindered by their L2 vocabulary knowledge. Thus, it is very important for them to learn high frequency words, academic words, and other words which occur in academic contexts in order to comprehend and communicate effectively, fluently, and accurately.

Research has shown that in view of the enormous vocabulary that students have to use in EAP, learning would be enhanced if students were provided with the list of frequent vocabulary found in academic text. These lists have been thought to

provide the vocabulary necessary to function in academic contexts (Coxhead and Nation, 2001). In 2000, Coxhead built the academic word list (AWL, please refer to *appendix 1*), which consists of 570 word families occurring frequently in written texts across a range of university disciplines. It has been considered to be very important and a useful learning goal for learners of English for academic purposes (Nation, 2001). The AWL has also been very useful and influential in recent years in the teaching and testing of English for academic purposes as a reference list for the sub-technical vocabulary that students are assumed to need in undertaking university studies through the medium of English (Nation, 2001; Read, 2007). Thus, the merits of the AWL have been highlighted by researchers, teachers, and students.

However, there are different reflections on AWL recently, in terms of field specific discipline. Researchers (Chen & Ge, 2007; Martinez, Beck & Panza, 2009) have shown that AWL can be classified as general academic vocabulary list. There is therefore a necessity for developing discipline specific academic wordlists to meet the needs of non-native English speakers (NNS) and writers who must read and publish articles in English. Likewise, there are other important words in a specific field that deserve further attention (Vongpumivitch, Huang & Chang, 2009).

In the globalizing world, communication across cultures and languages in all era of life has become more significant than ever before. Recently, Islamic scholarly writings are not only dependent on Arabic language but also have a clear priority, the urgent need to write papers in English. With the global focus on Islam, Arabic language has influenced many languages of the world. In spite of this, English is still the global language used not only for oral communication but also for academic writing. English is now being widely recognized as a *lingua franca* in many spheres of discourse. To date, the vast majority of academic, scientific and technical writings are

being published in the English medium. Thus, Islamic academic journals have begun publishing research papers in English. Therefore, the study of lexical items used in such academic articles has received great attention recently due to the increasing demand for instruction by NNS writers in academic settings.

The vocabulary that is widely used in this kind of articles can be classified as generic, because they functioned uniformly in all the disciplines such as science, social science or human science. On the contrary, they can also be viewed as specific, since they are all approached from only one source, that is, Islamic. Thus, on this account, there is a need to study the frequency of the AWL words that are used in the field of Islamic academic disciplines. Since the research article constitutes a particular genre, it can be assumed that the Islamic research articles corpus should contain a particular academic lexis, and that it might be fruitful to examine which words are particularly or unexpectedly frequent in this corpus.

## **1.2 RESEARCH PROBLEM**

One of the most complicated aspects of vocabulary learning and teaching in English for academic context is making principled decision about which words are worth focusing on during teaching and learning (Coxhead, 2000). As is known in academic settings, English for Specific Purposes (ESP) students do not have big problems with technical vocabulary because these words are usually the focus of discussions in the classroom or are considered enough to deserve highlighting in the textbooks (Stevens, 1973). It must be noted, however, that one of the most robust language phenomena of ESP learning in the academic field is known to be the acquisition of ‘academic’ vocabulary, or semi-technical vocabulary (Fuentes 2001; Li & Pemberton, 1994; Shaw, 1991; Thurstun & Candlin, 1998). Thus, even though ‘academic’ terms

are common across academic disciplines, they may cause problems for learners. That is, they are neither sufficiently frequent in the language as a whole to be learnt implicitly nor part of the technical lexicon which is likely to be explicitly taught as part of the subject courses (Nation, 2001).

Therefore, Coxhead (2000) has built the AWL, by compiling an academic corpus from Law, Arts, Commerce, and Science. Furthermore, Coxhead and Nation (2001) reported that the AWL words play a significant role in academic texts. It provided coverage around 8.5-10% of the words of an academic text, and combined with the General Service List (GSL, please refer appendix 2) (West, 1953), which was built with the most widely used, frequent words of the language, was expected to account for around 90% of any academic text.

In spite of its important coverage, some major issues have surfaced with regard to the validity and usefulness of AWL word frequency counts and the subsequent creation of high frequency word lists. These problems have caused researchers to come up with widely varying numbers and results concerning how many words ESL learners need to know for adequate comprehension and specific levels of proficiency (Nation, 2006).

As a result, Martinez et al., (2009) demonstrated that there is a need to produce field-specific academic word lists, as well as the specific academic vocabulary that needs to be taught at specific contexts. Moreover, recent studies on academic vocabulary using different corpora (Chen & Ge, 2007; Martinez et al., 2009) investigated the usefulness of the AWL in ESP courses with the claim that the list offers a general academic vocabulary and the lexical differences may exist across distinct disciplines. The bottom line in this case is that there still exists a remarkable imbalance between the frequencies of word occurrence in different disciplines. In

particular, the use of corpora has brought to light various trends in the frequency counts. For example, certain words occur with great frequency in a given text while others may occur only once or not at all. Thus, the frequency of the words selected for the AWL may be unbalanced when the perspective of subject discipline is taken into consideration. For example, a word which occurred 100 times in engineering texts might only have appeared 10 times in Islamic texts. Accordingly, it can be assumed, in relation to academic words, that not all the words in the list would be important for learners with highly specific needs.

It must also be noted that perhaps the most significant issue that might affect, among the various counts and lists, is the lack of agreement about what to count as a word, lexical item, or vocabulary unit. This may seem simple to answer, but there are many different individual and subjective decisions as to whether or not they should contain words from the source text or much related with writers' concept of word knowledge. It would differ from person to person. For example, words can be counted as individual forms, as lemmas, as word families, or in word collocations. The different concept of a word used as the unit of measurement in a frequency count will cause various results in the stages of developing word lists.

Thus, the aim of this study is to explore the frequency of the AWL words that are used in the field of Islamic academic research articles (IARA). The reason why this research is conducted is due to AWL words' central role in Islamic academic research articles. To the researchers' knowledge, research has yet to be conducted using the corpus of articles that have been written from the Islamic perspective. Hence, to fill this gap and to gain a better understanding of the performance of the AWL words in this area and to develop a reliable lexical syllabus for ESP students, the researcher conducted this corpus-based lexical study on the word families.

Furthermore, the researcher provided a selection of items from the AWL that is specifically useful to write in the specific genre and field of Islamic research papers. She also pointed out the need to include other lexical items that do not occur in the AWL but are used with academic function in this corpus. These results are expected to further contribute to the argument in favour of specificity in ESP courses.

### **1.3 OBJECTIVES OF THE RESEARCH**

This study embarks on the following objectives:

1. To examine the frequency, coverage and distributions of AWL word families in the Islamic academic research article corpus.
2. To identify the other high frequency items that do not occur in the AWL, but are used with academic function in the corpus of Islamic academic research articles.

### **1.4 RESEARCH QUESTIONS**

In order to find out the frequency, coverage and distribution of AWL in Islamic academic research articles and to see how Islamic academic English words represent and behave from GSL (West, 1953) and AWL (Coxhead, 2000) words, the following main research questions are asked:

1. What are the frequency, coverage and distribution of Coxhead's AWL word families in the Islamic academic research articles corpus?
2. Is there any lexis specific to Islamic academic English?
3. Are there any lexical items occurring more frequently and uniformly across a wide range of Islamic academic research articles but are not among the GSL and AWL?

## **1.5 SIGNIFICANCE OF THE RESEARCH**

The study is significant as it aims to explore the AWL words that are used in Islamic academic research articles. The significance of this study includes the following:

- The outcome of this research is expected to build a foundation for developing a reliable lexical syllabus by helping students in this area to focus on more useful vocabulary items, in order to meet the ESP course goals and learning objectives in the acquisition of academic vocabulary.
- It will be useful for text and dictionary writers in producing appropriate Islamic academic books, texts as well as dictionaries for ESP students.
- It also helps teachers and non-native English speaking students to choose more suitable teaching and learning materials in Islamic academic context.
- The expected results will lead to the development of discipline-specific academic wordlist to meet the needs of non-native English writers who must read and publish articles in English.

## **1.6 CONCEPTUAL DEFINITION OF TERMS**

The following are conceptual terms used in this study:

- *Types* are defined as single word forms. We can count the words in the sentence 'It is very difficult to read it quickly' in another way. If we see the same word again, we do not count it again. So the sentence of eight tokens consists of seven different words or 'types' (Nation, 2001:7).
- Tokens are defined as the number of occurrences of each 'type'. One way is simply to count every word form in spoken or written text, and if the same word form occurs more than once, then each occurrence of it is counted. So the sentence 'It is very difficult to read it quickly' would contain eight words, even

though two of them are the same word form, *it*. Words which are determined in this way are called ‘tokens’ or sometimes ‘running words’ (Nation 2001:7).

- A *word family* is defined as a collection of formally and semantically related word types (Baur & Nation, 1993). “A word family consists of a headword, its inflected forms, and its closely related derived forms” (Nation, 2001:8). Coxhead (2000:218) defined “*a word family*” as a stem plus all closely related affixed forms, as defined by Level 6 of Baur and Nation’s (1993) scale. The level 6 definition of affix includes all inflections and the most frequent, productive, and regular prefixes and suffixes. It includes only affixes that can be added to stems that can stand as free forms. (e.g., *specify* and *special* are not in the same word family because *spec* is not a free form).” For example, a word family from the AWL is illustrated in the following: *assign, assigns, assigned, assigning, reassign, reassigns, reassigned, reassigning, unassign, unassigns, unassigned, unassigning, assignment, assignments, reassignment, and reassignment*. These words are all the word family members of the same word family of **assign**
- *Academic vocabulary* has been called as academic vocabulary (Martin, 1976), semi-technical vocabulary (Farrell,1990), or sub-technical vocabulary (Cowan, 1974), and is viewed as “formal, context-independent words with a high-frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses; words with high-frequency across scientific disciplines” (Farrell, 1990, p. 11).
- A *Corpus* in this study refers to a principled collection of research articles held on a computer and by hand. They are stored electronically, and processed by a

computer program. The samples come from only one source, that is, Islamic academic journals.

- *Islamic academic research article* in this study refers to a paper or an article in an academic work that is usually published in an academic journal or is peer reviewed by one or more *referees* (who are academics in the same field) in order to check that the content of the paper is suitable for publication in the journal. It mostly contains original research results and would fall under any categories of research paper that are natural sciences, social sciences or humanities with special reference to Islam.

## **1.7 ORGANIZATION OF THE CHAPTERS**

This research has the following sequences of five chapters. An introduction of the study is in the beginning chapter which provides the background, purpose, problems, significance of the research and research questions. Chapter 2, the review of literature, provides relevant concepts, introduces related topics and findings from previous studies that support those concepts. Chapter 3 deals with research procedures and examines the data collected through article corpus needs. It discusses in detail the steps that have been followed before arriving at the final procedures. Chapter 4 presents the data according to the classification of the research objectives. It further analyzes the data to the extent that supports the research questions. All the high frequency words in the corpus were categorized and analyzed in order to show the frequency, coverage and distributions of AWL in this study. Finally, the last chapter provides the conclusion, recommendations and suggestions for further research.

## **CHAPTER TWO**

### **LITERATURE REVIEW**

#### **2.1 INTRODUCTION**

This chapter reviews the empirical research and theories in the area of Corpus linguistics, especially, corpus-based studies on frequency vocabulary in the academic context. The objective of this review is to build a foundation for a theoretical framework for this study. The following section presents the correlations between lexical knowledge and L2 proficiency in ESP learners in academic settings. It also discusses briefly the levels and categories of vocabulary in academic context and some of its coverage and distributions. The chapter then reviews some previous studies on academic words and discusses briefly Coxhead's (2000) AWL words and stresses its important role in academic life. It also reviews other studies on AWL words in different fields and points out the issues of its usefulness and coverage in relation to the frequency. Finally, the chapter summarizes the main points raised in the review.

#### **2.2 CORPUS LINGUISTICS**

In order to learn about word usage in a natural language, linguists have developed computer-based methods of examining authentic text. They have been able to collect large samples of language corpora, and analyze them by using sophisticated programs to produce frequency counts of word usage in both spoken and written texts. These language corpora are useful to language teachers in achieving their goals in teaching, and to language learners in achieving their goals in language learning. Today, corpus

linguistics offers some of the most powerful new procedures for the analysis of language, and the impact of this dynamic and expanding sub-discipline is making itself felt in many areas of language study (Wynne, 2005). A corpus is defined as a body of text that is representative of a particular variety of language and is stored on a computer. Originally done by hand, corpora are now largely derived by an automated process. The samples can come from anywhere the language is used in speech and in writing. It helps learners to understand more about the language and see how people use it when they speak and when they write.

For decades, corpus linguistics has been defined variously. McEnery, Xiao and Tono, note that as “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” (2006: 7-8); they therefore conclude that corpus linguistics is a methodology. Therefore, they claimed that corpus linguistics is a methodology. Corpus linguistics has also defined as a methodology in Meyer (2002), and as “an approach or a methodology for studying language use” in Bowker & Pearson (2002: 9). Thus, Corpus linguistics is a methodology which can be described as a study of authentic language on samples of ‘real life’ language use via a *corpus* (McEnery & Wilson, 1996 and 2001). Besides, Teubert (2005: 4) asserted that “Corpus linguistics is not in itself a method: many different methods are used in processing and analysing corpus data. It is rather an insistence on working only with real language data taken from the discourse in a principled way and compiled into a corpus.”

Reppen (2002) summarized the four essential characteristics of corpus linguistics as follows:

- a. It is empirical, analyzing the actual patterns of use in natural texts.

- b. It utilizes a large and principled collection of natural texts, known as a ‘corpus,’ as the basis for analysis.
- c. It makes extensive use of computers for analysis, using both automatic and interactive techniques.
- d. It depends on both quantitative and qualitative analytical techniques

Biber, Conrad and Reppen (1998) said that studying language can be divided into two main areas: studies of structure and studies of language use. From language use perspective, researchers can investigate how speakers and writers exploit the resources of their language rather than looking at what is theoretically possible in a language. In other words, researchers study the actual language used in naturally occurring texts. The goal of corpus-based investigations is not simply to report quantitative findings, but explore the importance of these findings for learning about the patterns of language use. That is, it is essential to include both qualitative and functional interpretations of quantitative patterns.

Sinclair (2005: 1-16) proposed 10 basic principles to construct a well-designed corpus as stated below:

1. The contents of a corpus should be selected without regard for the language they contain, but according to their communicative function in the community in which they arise.
2. Corpus builders should strive to make their corpus as representative as possible of the language from which it is chosen.
3. Only those components of corpora which have been designed to be independently contrastive should be contrasted.
4. Criteria for determining the structure of a corpus should be small in number, clearly separate from each other, and efficient as a group in