Language Tension, Terminology Variation and Terminology Policy in the Arabic-Speaking North African Countries: An Alternative Approach to Terminology Practice

A dissertation submitted to Kent State University
in partial fulfilment of the requirements for the
degree of Doctor of Philosophy

by
Fawzi Younis Hamed

Kent State University
December, 2014
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. VII
LIST OF FIGURES ................................................................................................................ IX
LIST OF MAPS ...................................................................................................................... X
LIST OF ABBREVIATIONS ..................................................................................................... X
ARABIC TRANSLITERATION SYSTEM ............................................................................... XII
DEDICATION ....................................................................................................................... XIII
ACKNOWLEDGMENTS ......................................................................................................... XIV

## CHAPTER ONE: OVERVIEW

1.1 Introduction .................................................................................................................... 2
1.2 Research Problem .......................................................................................................... 5
1.3 Research Questions and Hypotheses ............................................................................ 7
1.4 Significance of the Study .............................................................................................. 9
1.5 Structure of the Study .................................................................................................. 11

## CHAPTER TWO: HISTORICAL BACKGROUND AND SOCIOLINGUISTIC EVOLUTION IN THE ANAC

2.1 Introduction .................................................................................................................... 15
2.2 Arabic as a Colonial Language ...................................................................................... 16
2.3 Arabic Diversity: Classical Arabic CA vs. Modern Standard MSA ...................... 19
2.4 Arabic Dialects ............................................................................................................. 23
2.5 Arabic Diglossia ........................................................................................................... 29
2.6 The Role and Status of French .................................................................................... 32
2.7 The Role and Status of English ................................................................................... 37
2.8 Indigenous Languages ................................................................................................. 42
2.9 Language Competition and Language Attitude ......................................................... 46
2.9 Summary and Conclusion ......................................................................................... 52
CHAPTER THREE: LP AND TP: THEORETICAL AND COMPARATIVE BACKGROUND

3.1 Introduction .......................................................... 56

3.2 Language Planning .................................................... 54

3.3 Language Policy vs. Language Planning ......................... 59

3.4 Language Planning Typology .................................... 60

4.5 Terminology Planning ............................................. 62

3.6 Who Plans Terminology? ............................................. 65

   3.6.1 Top-down vs. Bottom-up Model ......................... 65

   3.6.2 Integrative Model ............................................. 70

3.7 Term Formation ..................................................... 71

   3.7.1 Domain Loss in the Field of Computers and the Internet .... 74

   3.7.2 Methods of Term Formation ................................ 76

      3.7.2.1 Creating New Terms .................................. 77

         3.7.2.1.1 Derivation ........................................ 77

         3.7.2.1.2 Compounding ...................................... 81

         3.7.2.1.3 Abbreviation ...................................... 83

      3.7.2.2 Translingual Borrowing and Loan Translation ........... 85

         3.7.2.2.1 Loan Words (direct loan) ....................... 85

         3.7.2.2.2 Loan Translation ................................ 88

      3.7.2.3 Arabisation .............................................. 90

3.8 Terminology Practice and Official Terminology Policy in the AW .... 92

3.9 Summary and Conclusion ......................................... 97

CHAPTER FOUR: METHODOLOGY OF THE STUDY .................. 99

4.1 Introduction ........................................................ 100

4.2 Corpus Development and Terminology Extraction ............... 101

   4.2.1 General ...................................................... 101
CHAPTER FIVE: PATTERNS OF REGIONAL VARIATIONS IN ARABIC TERMINOLOGY

5.1 Introduction
6.3.5.3 Standardization .......................................................... 224
6.3.6 Pedagogical Aspect ....................................................... 232
6.4 Conclusions .................................................................... 239
6.5 Limitations and Future Research ........................................ 247

Appendix I: Websites from Which Corpora Were Collected .......... 249
Appendix II: Stop List of Arabic Words and Their English Equivalents ...
Appendix III: Corpus One Term list .......................................... 251
Appendix IV: Corpus Two Term list .......................................... 252

WORKS CITED ..................................................................... 253

LIST OF TABLES
Table 2.1  Regional dialects of Arabic ........................................ 25
Table 2.2  Changes in the phoneme qaaf in regional dialects ........... 26
Table 2.3  Distribution of languages by countries .......................... 48
Table 3.1  Derivation methods ................................................... 76
Table 3.2  Terms of Italian origin in Libyan Arabic ....................... 84
Table 4.1  Arabic corpora ......................................................... 108
Table 4.2  Hindi and Arabic numerals ........................................ 109
Table 4.3  Different shapes of Arabic character ‘hah’ ..................... 111
Table 4.4  Different shapes of Arabic character ‘hah’ in words .......... 111
Table 4.5  Arabic derivation (verbs) ........................................... 113
Table 4.6  Arabic derivation (nouns) ........................................... 113
Table 4.7  Summary of Arabic short vowel signs .......................... 115
Table 4.8  Ambiguous word list from the Buckwalter corpus .......... 116
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Arabic categories (diglossia)</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>Arabic categories (triglossia)</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Types of LP</td>
<td>59</td>
</tr>
<tr>
<td>3.2</td>
<td>The relationship between LP and TP</td>
<td>62</td>
</tr>
<tr>
<td>3.3</td>
<td>Top-down model</td>
<td>65</td>
</tr>
<tr>
<td>3.4</td>
<td>Bottom-up model</td>
<td>66</td>
</tr>
<tr>
<td>3.5</td>
<td>Integrative model</td>
<td>69</td>
</tr>
<tr>
<td>3.6</td>
<td>Stages of the terminology development process</td>
<td>71</td>
</tr>
<tr>
<td>4.1</td>
<td>Qur’anic Arabic Corpus Syntactical Treebank</td>
<td>106</td>
</tr>
<tr>
<td>4.2</td>
<td>Distortions in Arabic text</td>
<td>110</td>
</tr>
<tr>
<td>4.3</td>
<td>Properly displayed Arabic text</td>
<td>110</td>
</tr>
<tr>
<td>4.4</td>
<td>(ṣ) and in Arabic corpora</td>
<td>119</td>
</tr>
<tr>
<td>4.5</td>
<td>Wordlist as displayed by WST</td>
<td>122</td>
</tr>
<tr>
<td>4.6</td>
<td>Arabic concordance as displayed by WST</td>
<td>124</td>
</tr>
<tr>
<td>4.7</td>
<td>Word list before the generation of function words stop list</td>
<td>135</td>
</tr>
<tr>
<td>4.8</td>
<td>Lemmatization of different inflectional words</td>
<td>137</td>
</tr>
<tr>
<td>4.9</td>
<td>Arabic concordance using WST</td>
<td>139</td>
</tr>
<tr>
<td>5.1</td>
<td>Orthographic variation</td>
<td>150</td>
</tr>
<tr>
<td>5.2</td>
<td>Borrowed brand names written in both English and Arabic</td>
<td>159</td>
</tr>
<tr>
<td>5.3</td>
<td>(ṣ) waw and as a bound token</td>
<td>178</td>
</tr>
<tr>
<td>5.4</td>
<td>(ṣ) waw and as a separate token</td>
<td>178</td>
</tr>
<tr>
<td>5.5</td>
<td>Use of foreign-language abbreviations in Arabic texts</td>
<td>188</td>
</tr>
<tr>
<td>5.6</td>
<td>Excessive use of French in Arabic texts</td>
<td>192</td>
</tr>
</tbody>
</table>
LIST OF MAPS

Map 1.1 Arabic-North African Countries (ANAC) .............................. 4
Map 2.1 Arabic regional dialects ..................................................... 24
Map 2.2 Major Libyan Arabic dialects ............................................. 27
Map 2.3 Major Libyan Amazigh groups ............................................ 27
Map 2.4 Geographical distribution of Berbers in ANAC ..................... 44

LIST OF ABBREVIATIONS

ANAC Arabic-North African Countries .............................................
AW Arab World .................................................................
BiDi A Mixture of Right-to-left and Left-to-right Direction .................
C1 Morocco, Algeria and Tunisia Corpus ......................................
C2 Egypt and Libya Corpus ....................................................
CA Classical Arabic ..............................................................
CBA Coordination Bureau of Arabization ......................................
CP Corpus Planning ..............................................................
CP Communication Planning ....................................................
DEAP Declaration of the Establishment of the Authority of People ....
H High Language Variety .......................................................
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSA</td>
<td>Modern Standard Arabic</td>
</tr>
<tr>
<td>ISO</td>
<td>The International Organization for Standardization</td>
</tr>
<tr>
<td>L</td>
<td>Low Language Variety</td>
</tr>
<tr>
<td>LP</td>
<td>Language Planning</td>
</tr>
<tr>
<td>LSP</td>
<td>Language for Special Purposes</td>
</tr>
<tr>
<td>LTR</td>
<td>Left-to-right</td>
</tr>
<tr>
<td>OCR</td>
<td>Optical Character Recognition</td>
</tr>
<tr>
<td>RTL</td>
<td>Right-to-left</td>
</tr>
<tr>
<td>SP</td>
<td>Status Planning</td>
</tr>
<tr>
<td>ST</td>
<td>Source Text</td>
</tr>
<tr>
<td>SVO</td>
<td>Subject Verb Object</td>
</tr>
<tr>
<td>TM</td>
<td>Translation Memory</td>
</tr>
<tr>
<td>TP</td>
<td>Terminology Planning</td>
</tr>
<tr>
<td>TT</td>
<td>Target Text</td>
</tr>
<tr>
<td>TTR</td>
<td>Type-token Ratio</td>
</tr>
<tr>
<td>UNESCO</td>
<td>United Nations Education, Scientific and Cultural Organization</td>
</tr>
<tr>
<td>VSO</td>
<td>Verb Subject Object</td>
</tr>
<tr>
<td>WST</td>
<td>WordSmith Corpus Tool</td>
</tr>
</tbody>
</table>
ARABIC TRANSLITERATION SYSTEM

Adopted from the Library of Congress transliteration system for Arabic script, the following transliteration system has been consistently utilized throughout the present dissertation whenever an Arabic term or expression is quoted. The following table explains the Arabic transliteration system for consonants and vowels.

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Latin</th>
<th>Arabic</th>
<th>Latin</th>
</tr>
</thead>
<tbody>
<tr>
<td>أ</td>
<td>a</td>
<td>ض</td>
<td>dh</td>
</tr>
<tr>
<td>ب</td>
<td>b</td>
<td>ط</td>
<td>i</td>
</tr>
<tr>
<td>ت</td>
<td>t</td>
<td>ض</td>
<td>z</td>
</tr>
<tr>
<td>ث</td>
<td>t</td>
<td>ع</td>
<td></td>
</tr>
<tr>
<td>ج</td>
<td>h</td>
<td>غ</td>
<td>gh</td>
</tr>
<tr>
<td>ح</td>
<td>h</td>
<td>ف</td>
<td>f</td>
</tr>
<tr>
<td>خ</td>
<td>kh</td>
<td>ق</td>
<td>q</td>
</tr>
<tr>
<td>د</td>
<td>d</td>
<td>ك</td>
<td>k</td>
</tr>
<tr>
<td>ذ</td>
<td>d</td>
<td>ل</td>
<td>l</td>
</tr>
<tr>
<td>ر</td>
<td>r</td>
<td>م</td>
<td>m</td>
</tr>
<tr>
<td>ز</td>
<td>z</td>
<td>ن</td>
<td>n</td>
</tr>
<tr>
<td>س</td>
<td>s</td>
<td>ه</td>
<td>h</td>
</tr>
<tr>
<td>ش</td>
<td>š</td>
<td>و</td>
<td>w</td>
</tr>
<tr>
<td>ص</td>
<td>š</td>
<td>ي</td>
<td>y</td>
</tr>
</tbody>
</table>

Short/long vowels and case endings

| | werde | | |
|---|---|---|
| أ | ā | ي | ī |
| و | ū | َ | a |
| َ | u | َ | an |
DEDICATION

This dissertation is dedicated to my wonderful wife, Ghada. She is the love of my life, and the strength and support in this endeavour. I also dedicate this to my loving children Tasneem, Rayan, Ansam, Zakaria and Ellan, who are my constant source of strength and reminders of who I am really fighting this battle for. I have a great love for you all.
ACKNOWLEDGMENTS

Many people helped in making this dissertation a success. I take the opportunity to express my gratitude to all those who directly or indirectly have contributed in the completion of this dissertation.

First and foremost, I would like to express my sincere apologies to all my children, my wife and family who suffered during the writing of this PhD marathon. I would like to thank my wife Ghada for her unwavering support and love she has shown during the past four years. This dissertation would not be possible without her encouragement, patience, strength and love throughout the years. My second great gratitude goes to my dissertation supervisor professor Sue Ellen Wright for her invaluable guidance, useful comments and remarks as well as for her support and encouragement through the writing process of this dissertation. Also, I would like to thank my mother for her prayers and her spiritual support throughout my life. Thanks to my colleagues and faculty members of the Institute of Applied Linguistics at Kent State University for their support over the past years, I benefited greatly from their knowledge and experience during courses and projects. Last but not the least, my deepest gratitude goes to the dissertation committee members Professors Judy Wakabayashi and Gregory M. Shreve for their valuable comments and suggestions, as well as Professor Joshua Stacher of the Department of Political Science, who has served as my outside reader.

Fawzi Hamed
Fall 2014, Kent, Ohio
CHAPTER ONE

OVERVIEW
1.1 Introduction

The Arab-North African countries (ANAC; i.e., Egypt, Libya, Tunisia, Algeria and Morocco, see Map 1.1) have long been mistakenly characterized by the omnipresence of Arabic, or of so-called Modern Standard Arabic (MSA). The North African Arabic region is considered today to be perhaps the most multilingual region in the Arab World (AW). This consists, partly due to the hegemony of foreign languages, mainly French and English. It might be also attributed to the regional variations that the Arabic language exhibits, to the point that these variants are somewhat incomprehensible to users from different dialect areas, as well as the widespread use of indigenous languages.

This study is concerned with examining this multilingual situation in the region and its social, scientific and political impact on terminology policy (TP). It is based on the hypothesis that two factors are behind the phenomenon of term variation and term formation. The first is the influence of foreign languages, mainly English and French, and the second is the influence of local dialects. The numerous regional dialects and the hegemony of foreign languages are likely to affect technical term formation and use. To test these hypotheses, the study addresses the current issues in implementing an effective TP in the ANAC region in particular and in the AW in general. This investigation is accompanied by an examination of terminological variation in context. Terminological variation can be basically defined as the use of alternative linguistic forms that express one specialized concept. It includes many different phenomena (e.g., orthographic variation, lexical variation, morphological variation, reduction, permutation, and short forms). The study uses a corpus-based approach to terminology, which opens up the possibility of gathering both conceptual and linguistic information as well as usage information about the terminological units (Sager 133). This approach also supports the observation of term variation and allows comparison of Arabic texts from the specialized subject field of
computers and the Internet so as to extract valuable information about how TP can be effectively implemented for terminology standardization purposes.

The field of computers and the Internet is particularly relevant, because it presents a high level of difficulty for translators for the following reasons: first and foremost, terminological variation is a widespread phenomenon in specialized discourse (see Freixa 2001), especially in the dynamic discourse of computers and the Internet, where terms are continually born to express different newly developed concepts. Second, the field of information technology in general has been introduced to the region since the colonialist era from outside, thus it is likely to employ novel terminologies. Third, in the field of computers and the Internet, content tends to be relatively dynamic and complicated, because not only are the translator and the technical writer expected to be well-versed in the subject-field terminology of the two languages involved, but they need also to master the scientific realities of the text content being translated as well. The study concludes with how language and terminology policy could improve the current terminology situation by establishing the necessary foundation for a systematic methodological framework for formulating and implementing a TP, as well as by defining the common factors that have impeded progress in creating an effective TP with a view to standardizing and harmonizing Arabic computer and Internet terminology throughout the region.
Although the entire ANAC is Arabic-speaking and MSA is the only official language, the current sociolinguistic situation is vastly more complex and dynamic than in other Arabic regions such as the Arabian Gulf and the Middle East, simply due to the linguistic diversity of the region. The conflict of prestige and power between the three colonial languages in the region, namely Arabic, French, and English, is an increasingly noted phenomenon. The continued hegemony of French as a tool for modernization and development and the growing demand for English as a medium for modern science, technology and economic development, as well as the growing consensus among Arabic scholars that Modern Standard Arabic MSA is the only acceptable solution for an effective scientific language, have resulted in the phenomenon of discrepancy and inconsistency of terminology across the region.

In such a multilingual situation, where different languages are in contact, translators and technical writers are apt to accept or adopt at least one of two Arabic equivalents or terms for some foreign technical terms in accordance with their dominant foreign languages, either French or English. French-competent translators might adopt one Arabic equivalent
from French and English-competent translators might adopt a different Arabic equivalent from English. The adoption of foreign terms is also seen as a legitimate process, especially when Arabic does not have an equivalent for the term used. Jakobson supports this notion, as he says, “Whenever there is deficiency, terminology may be qualified and amplified by loanwords or loan-translations” (234). A loan word or term is a word that is accepted directly into a target language (TL) without significant change in sound or spelling. A loan translation involves the translation of the morphemic elements making up a multi-morpheme source language (SL) term into the TL. Translators are also apt to adopt local dialect roots for coining new terms if no other alternatives are found in MSA. Other studies have indicated that colloquial terms are often the only option that exists for certain discourses, for example, agriculture and food production (see Wilmsen and Youssef 2009). I have also demonstrated later in this study (Chapter Three) that Libyan colloquial terms are the only ones that exist for Libyan users in the arena of automotive discourse.

Thus, it is up to the translators or the technical writers who face the problem of not finding an equivalent in MSA for a particular foreign term to choose the most suitable way to render it by coming up with their own equivalents, a task which they often pursue without following any systematic criteria for forming terms in Arabic. This terminological phenomenon will be identified and viewed in its wider context in Chapter Three.

1.2 Research Problem

Term formation should not be a random or arbitrary process, but rather

a conscious activity which differs from the arbitrariness of general word formation because term formation requires a greater awareness of pre-existing patterns and models, and of the social responsibility for facilitating communication and transmitting knowledge. (Sager 25)
Term formation takes different forms (derivation, compounding, borrowing, blending, abbreviation and so forth) and occurs in different environments, e.g., governmental institutions, professional organizations, universities, and among individual translators and writers. Ideally, however, subject field experts working as terminologists and translators play a crucial role because they have the insight and expertise into the problems of their domain. The problem with Arabic terminology formation in general is that term formation and production is scattered and unsystematic, which leads to a variety and multiplicity of synonymous Arabic terms for the same concepts. This variation is most acutely felt in the arena of specialized terminology, such as computer and Internet terminology. A classical example is the English term ‘computer,’ which is translated into Arabic as hasub الحصوب, Alhasib الحاسب, hasibali الحاسب الآلي, rattaba الرتبة or kombiuter الكمبيوتر. Rendering the concept as hasub الحصوب with a short vowel or hasoob الحاسوب with a long vowel may be found in North African dialects, particularly Tunisia and Libya. The loan translation hasibali الحاسب الآلي (literally meaning ‘computing machine’) is seen in writing in Egypt and the Middle East, whereas rattaba الرتبة (literally meaning ‘ordering things’) is only seen in Algeria and Morocco’s dialects, where French is the dominant colonial language (for more discussion about this example in particular see 3.7 Term Formation). In general, there are two different linguistic sub-regions with regard to foreign language dominance: a French orientation in Morocco, Algeria, Tunisia and Mauritania, and an English orientation in Libya, Egypt and almost all of the rest of the AW. This means that there are two different foreign language sources for terminology, and the pace and scope of Arabisation has depended largely on which of these languages is more dominant. In many cases events, Arabic terms are bound to be influenced by one or the other source language, either French

1 Arabisation is the introduction of a foreign term into Arabic. It is the second one of the main methods for creating neologisms when developing Arabic terminology. Arabization, on the other hand, is the process of imposition and usage of Arabic language in place of some other languages (see Chapter Three for a distinction between the two terms).
or English. Consequently, terminological variation may occur as a result of loanwords from these two different language sources.

These loanwords sometimes do not convey all the meanings and usages of certain concepts. Thus their Arabic equivalents will not be adequate. The translator, therefore, has to determine whether or not Arabic linguistic and terminological bodies, such as the Academy of Arabic Language, have translated those terms, as they are supposed to have the insight and expertise to produce translations and equivalent lexical items for most of the new and up-to-date terms and expressions in different fields of knowledge. The problem, however, is that language academies working in various parts of the ANAC typically respond slowly to the flood of new terminology, especially in the area of information technology.

1.3 Research Questions and Hypotheses

1. The study proposes that traditional approaches to Arabic TP in general are inadequate in addressing terminology problems in such a multilingual context, as well as in dealing with the flood of new terminology, particularly in specialized and rapidly changing areas such as computers and the Internet. The study, therefore, explores the current implementation of TP in the ANAC region in order to determine how stakeholders (linguists, translators, LSPs [Language Service Providers]), private and public institutions, companies, governments, and so forth) in the region may create and implement a comprehensive and systematic formulation and implementation of TP. In order to answer this question the study defines the potential factors that have impeded progress toward an effective TP. The study also proposes an alternative integrative approach to terminology work. The approach argues that an adequate model of TP should accommodate a variety of stakeholders.
2. The study also proposes that one of the many reasons for TP inadequacy in the ANAC is the dialectal differences among the countries of the region and the hegemony of foreign languages, which in turn have led to variation in communicating specialized information. Another fundamental question arises as to whether this ongoing language diversity in the ANAC is likely to show a large degree of terminological variation. In particular I will investigate whether dialectology and colonialism have any impact on term formation and use. I hypothesize that the numerous regional dialects and the hegemony of foreign languages are the main determiners for term selection and term formation in the ANAC. Texts used throughout the analysis are expected to reflect terminological variation in the arena of computers and the Internet based on foreign language preference as well as on the variety of Arabic equivalent options for a given term. The data have been analysed to test the hypothesis and thus to reveal terms that we can assume represent the same concept, but that may vary orthographically, lexically, morphologically, or syntactically. The steps that were taken in the process of data collection and the procedure of building and analyzing the corpora as well as the software that was used in the study are discussed in detail in Chapter Five.

While the question of regional variation within MSA both in its written and spoken forms and the influence of foreign languages on Arabic LP in general has been the focus of research for many decades (numerous studies such as Abu-Abssi, 1986; Alrabaa, 1986; Gully, 1993; Suleiman, 2003 have addressed this issue), little attention has been paid to the impact of this variation on terminology processes. Until now there are, as far as I know, no studies that explore the possible impact of the current multilingualism in the ANAC on term formation and use.
1.4 Significance of the Study

As Koffi (9) puts it, societies cannot achieve any significant level of modernization and development without systematic terminology use and treatment. Therefore, assessing the terminological variations in formal written MSA (the main focus of this study) is becoming a must. It is unquestionable that computers and the Internet are part of everyday life; they are used in universities, schools, homes, markets, and banking, both online and onsite. Hence, the use of computer terms is comprehensive and not limited to merely one social or professional group. Terms from this domain are widespread among specialists and non-specialists alike. Therefore, the need for consistent treatment (formation, standardization, dissemination and documentation) and use of computer and Internet terminology throughout countries, organizations, individuals, and various social strata to improve communication and integration between production and documentation is becoming indispensable for Arabic to function effectively (either internationally or locally) as a scientific and economic language.

Along with discussing the factors that have impeded progress in creating an effective TP, the study will provide language and TP makers in the region with theoretically sound and evidence-based recommendations that can be implemented by both governmental and non-governmental institutions, based on the analysis of the two study corpora. The recommendation are introduced in six aspects that may explain the major factors that have impeded progress in creating and implementing an effective TP in the ANAC region in particular and the AW in general.

In the age of globalization, consistent use of terminology can help organizations to ensure a uniform internal and external image and avoid unnecessary variants, corrections and miscommunication, resulting in substantial savings in time and money. A survey
conducted by SDL Trados in 2013 among companies in the localization and translation industry in Europe (1100 companies) asked respondents to look at the effect that terminology has on their business; 75% of the respondents thought that the biggest concern with inconsistent terminology in business was its negative impact on communication within and across organizations. Global organizations encourage the use of consistent terms throughout their websites, help files, documentation, and applications so as to improve communication. However, in the age of specialization where most scientific communication is conducted in English (Ibrahim 2009), many languages are now facing new challenges, and Arabic is no exception. Rapid advances and the hegemony of English and French as global languages have marginalized Arabic as a means of communication, especially in scientific and technological areas. Education, even at basic levels, is still conducted mostly in French or English, a particular problem in the more prestigious disciplines, such as medicine, engineering and information technology. This situation is fuelled by the belief among parents that there is an employment and educational advantage of learning these two languages instead of MSA. As far as terminology is concerned, the infusion of computer-related terminology into everyday language makes children quickly learn terms in accordance with language prestige and language preference. In such a scenario, it is becoming difficult to educate children in disciplines such as computers and the Internet where even basic concepts are constantly absorbed from foreign languages. Children are also at risk of losing their MSA vocabulary in specialized domains.

This study can be regarded as a modest contribution to countering this current trend in Arabic toward ‘domain loss,’ i.e., loss of the capacity in Arabic to generate new terms and expressions in certain technical and scientific fields such as computers and the Internet due to the hegemony of foreign languages and the insufficient mastering of Arabic MSA as a lingua franca for the majority of the population. MSA is in fact formally taught at
elementary and early secondary schools almost as a second language (Redouane 202). As a result, large sectors of the Arab public do not command MSA sufficiently. This fact has made it difficult for Arabic speakers to develop successful scientific and technical terminologies, and it has forced even experts to use foreign language terminologies for communication in scientific and technical domains. Many believe (e.g., Bakry 1999) that action is needed to protect Arabic language from the rush of foreign technical terms, particularly French and English. This study is important in that it exposes the widespread adoption of terms from foreign languages into Arabic and highlights the immediate need for the Arabisation of terms, especially in the arena of computers and Internet discourse where the foreign language effect is acutely felt.

1.5 Structure of the Study

This dissertation is composed of six chapters, including this introductory chapter. The structure of these chapters is as follows.

Chapter Two surveys the linguistic situation in the ANAC region. The chapter begins with a review of the history of the Arabic language and its development from the classical form into its modern form. It also describes the Arabic status of diglossia and triglossia. It briefly overviews the current status of French, English and the indigenous languages in the ANAC region and examines their role and relation with Arabic language.

Chapter Three provides terminological definitions in the field of LP and TP and discusses the key actors and elements involved in these two related fields. The chapter also reviews some critical approaches and contributions in LP and explores the relationship between LP and TP with the aim of locating TP in a LP framework. This chapter also investigates the applicability of different patterns (e.g., derivation, compounding, loan translation, Arabisation, abbreviation) put forward by Arabic linguists, translators, technical
writers, etc. in order to face the flood of new terms being created almost every day. The chapter finishes with a review of the terminological situation in the ANAC region.

Chapter Four gives a general overview of the methodology adopted in creating the two monolingual Arabic corpora of the special field of computers and the Internet with the aim of studying terminological variation in the Arabic language. The chapter also gives a brief definition of corpus linguistics and its types and provides an overview of some current Arabic corpora produced by different institutions and companies. It also discusses some of the common problems faced when developing a corpus for the Arabic language from written texts.

Chapter Five portrays the phenomenon of Arabic terminology variation through a comparison and analysis of the two corpora of texts relating to computers and the Internet generated from the two sides of the ANAC region (French-oriented countries and English-oriented countries). The focus of the analysis is primarily on orthographical, morphological, lexical and phonological variations. Acronyms and abbreviations are also discussed. The chapter also discusses both linguistic and non-linguistic factors that contribute to the current terminological diversity in the domain of computers and the Internet.

Chapter Six provides some conclusions about the research questions and describes some of the possible implications for TP and practice in the AW in general and in the ANAC region in particular. Six aspects that may explain the major factors that have impeded progress in creating and implementing an effective TP are identified, with a view to standardizing and harmonizing Arabic computer and Internet terminology throughout the AW. The limitations of the study are also noted in this chapter, and some suggestions for future research are made.
CHAPTER TWO

HISTORICAL BACKGROUND AND SOCIOLINGUISTIC EVOLUTION IN THE ANAC
2.1 Introduction

The ANAC region has experienced several documented layers of colonization since the Phoenicians occupied North Africa and established Carthage in 814 BC. As a result, it is considered today to be perhaps the most multilingual region in the AW in the sense that many languages and varieties are used—specifically MSA, colloquial Arabic varieties, French, Berber, and recently English. Therefore, the current multilingualism that characterizes this region of the AW is the ultimate outcome of that long colonization history. The argument here is that much of the current debate on Arabic language variety cannot make sense without connecting it to colonialism.² To put the current linguistic situation in context and to highlight the sociolinguistic differences and similarities between parts of the region, I will give a historical overview of the linguistic discourse of the ANAC.

This chapter is composed of seven sections. The first section reviews the introduction of the Arabic language to the region and discusses its development from classical form into modern form. It also describes the Arabic status of diglossia (the co-existence within a society of two quite different language varieties, each used in different social contexts.) Section Two begins with a brief overview of the current status of French and examines its role in the education system today. This section will focus more closely on Morocco, Tunisia and Algeria, as the linguistic situation there is dominated by French, and the relationship between French and other languages is relatively complex. Section Three discusses the emergence of the English language as another linguistic option for modern science and technology. The spread of indigenous languages throughout the region and the relation between them and Arabic will be discussed in Section Four. The aim here is to provide a socio-historical contextualization of the linguistic discourse and to understand the

² Ania Loomba in her book Colonialism/Postcolonialism provides a comprehensive definition for the term colonialism and makes a useful distinction between terms such as postcolonialism, imperialism and decolonization. However, the terms colonization and imperialism are still ill-defined as they often used interchangeably. Colonialism is also reproduced in its later sense with concepts such as globalization and Westernization. It emerged into a new concept of Western economic, cultural and even political hegemony.
current distribution of these languages in the ANAC. Section Five briefly defines the concept of language tension in the ANAC context, where Arabic vies with French, English and the indigenous languages. It also seeks to understand the ongoing spread and the changes in foreign language use and attitudes in the region in favour of English. Finally, Section Six concludes that due to this linguistic diversity and the tension among languages, the current sociolinguistic situation in ANAC is vastly more complex and more dynamic than in other Arabic regions, such as the Arabian Gulf and the Middle East.

2.2 Arabic as a Colonial Language

The appearance of the Arabic language in North Africa is inextricably intertwined with Arab conquests and the rise of Islam as a major religion in the region, which started with the conquest of Egypt in 637 under the rule of Omar Ebnl Al-khtab, the second Caliph after the death of the Prophet Mohammed. By the end of the tenth century, Arabic was widely spoken and written by all literate North Africans from Egypt to the Western borders of Morocco, displacing Berber languages, the oldest languages spoken by the indigenous people. Before the arrival of the Arabs, Berber languages, were the native languages for the majority of North Africa’s population. As Brenzinger (125) puts it, “Before the Arab invasion, the entire northern part of the African continent might have been a vast contiguous area of an Amazigh dialectal continuum in which neighboring varieties were mutually intelligible.” Hammoud (19) claims that besides Tamazight there were two other languages in use before the coming of Arabs in the seventh century. First, Latin was used as an administrative language until it was replaced by Arabic. Second, a hybrid combination of Latin, Greek and Semitic elements was developed in Carthage, a major urban center that evolved from the Phoenician colony and has existed in the region for nearly 3,000 years.

3 Different terms are used by different Berber groups to refer to the language they speak. For example, Algerian Berbers call their language Tazmat, while in Tunisia the local Berber language is usually referred to as Shelha. The term ‘Tamazight’ is used throughout this study to refer to all Berber languages, because this term is commonly used by many Berber linguists. The term ‘Amazigh’ is used in this study to refer to Berber people.
However, a recent study conducted by Moroccan History Television in 2012 supports the later claim. The Caspians, who came from East Africa (Ethiopia), inhabited the region thousands of years before the appearance of Berber in North Africa.

Indeed, the power of a language derives from the power of its speakers. “Language is made strong, prosperous, and widespread as long as its community is getting strong” (Altwaijri 15). As a matter of fact, the status of the Muslim nation during its historical phase (from the 7th century to the Mongolian invasion of Baghdad in the 11th century or what is called ‘the Islamic golden age’) gave Classical Arabic (CA) an early prominence and dominance. Early in the seventh century and not long after the conquests by the Arab armies from the Arabian Peninsula, the Arabs imposed the unity of Islam and the Arabic language, a concept based on the principle of the unity between the Muslims’ sacred text (the Qur’an) and Arabic. Muslim scholars have emphasized the untranslatability of the Qur’an and the use of Arabic language by non-Arabic Muslims in their daily prayers and for other religious purposes, such as Qur’an recitation. They have stressed the importance of the Qur’anic verses that explicitly state that only Arabic is the vehicle of the divine word, such as

(Qur’an 12: 2) {اَنَا أَنْزَلْنَاهُ ﻓِرْعَوْنَ ﺑَيْنَاءَ عَرَبِیَّ}{n-naa anzalnaahu qur’aanan ‘arabiyy-yan la’al-lakum ta’aqiluun}

“in-naa anzalnaahu qur’aanan ‘arabiyy-yan la’al-lakum ta’aqiluun”

“We have revealed the Qur’an in the Arabic tongue so that you may grow in understanding.”

The Prophet Mohammed himself repeatedly highlighted the notion of Arabic superiority and the inseparability of Arabic language and Islam. In one Hadith (Prophet’s tradition), cited in Suleiman (43), the Prophet is reported to have said that

The angel Gabriel descended from heaven and said to him: O Mohammed! All things have a master; Adam is the master of men, you are the master of Adam’s descendants …, the master of the trees is Lotus (Sider), the master of
master of birds is the eagle, the chief of months is Ramadan, the chief of weekdays is Friday, and Arabic is the master of speech.

Therefore, any attempt to reproduce the Qur’an in another language is futile. This conviction is still widespread among Muslims, i.e., “the doctrine that classical language as codified by the Arabic philologists, the language in which are written the sacred Qur’an and the classical works of Arabic literature, is unchangeable and is the only one to be used when writing” (Blau 7). As a result, non-Arabic-speaking Muslims, like the Amazigh who converted to Islam, began to utilize Arabic script. They had to learn Arabic so they could use it in their Qur’an recitations and daily prayers, salah. As the language of the Qur’an, CA became inseparable from Islam, with the result that Arabic spread widely in the region as part of the new religion and the only suitable tongue with which the masses of newly Muslim Amazigh could approach Allah. This superiority of Arabic and the inseparability of Arabic and Islam create a sense of identity among the Arabs that is centered on the Arabic language, thus hindering the acceptability of any foreign impact on the classical form of the Arabic language.

Since the seventh century and as the Arabs expanded their empire outside the Arabic peninsula (a period known as ‘the golden age of Arabic language and literature’), there has been continual contact between Arabic and Indo-European languages. After the spread of Islam in the seventh century and the imposition of Arabic as the only vehicle of the divine word as well as the contacts of Arabs with neighboring language communities (e.g., Spain, Portugal, France and England), many terms of Arabic origin were introduced to these communities, to the point that words of Arabic origin became so familiar that many European scholars in the eighteenth century ceased to regard the lexical wealth of Arabic, especially in math, science and philosophy, as a real source of richness of their languages (see Marzari
2006, 20-29). Examples include *cadi* 'قاضي', *cipher* 'صفر', *algebra* 'الجبر', *alcohol* 'الcohol'. It is estimated that there are more than four thousand words in Spanish that have Arabic origins. This influence on Spanish is considered the second largest after the influence of Latin (Quintana 699). Some linguists (e.g., Marzari 25) go so far as to point out that more than half of the Ottoman lexicon was of Arabic origin.

### 2.2.1 Arabic Diversity: Classical Arabic vs. Modern Standard Arabic

There are two varieties of Arabic: CA and MSA. CA is the language of the Qur’an, the Hadith, and the literature of the pre-Islamic and early Islamic periods (specifically, between the seventh and the ninth centuries). As the language of sacred text, CA is the high variety, and it still enjoys a great literary and religious tradition. It can be learned in schools and mosques alike. MSA, on the other hand, is the medium of contemporary learning and culture for Arabic societies and is considered to be a development of CA. It shares most of its syntax and morphology with CA. MSA is mainly used as a medium of communication only among educated Arabic speakers in conferences, gatherings of Arabic scholars, and for formal school education, specifically with respect to reading and writing. The reason for this limitation is the discrepancy between Arabic written and spoken forms. MSA is also used in printed media (newspapers and magazines) and as a spoken medium to various degrees in radio and television news.

MSA, in its modern form, was born in the period of the Abbasid Caliphate (750 – 1258) when Arab philosophers – some of them were translators⁴ – began to translate texts, particularly from Greek, but also from Chinese, Persian, Sanskrit, Sicilian and Spanish into Arabic to meet the increased research and scientific needs. Thus they modernized CA with respect to lexicon and phraseology by adding a large amount of vocabulary from these

---

⁴ Al-Kindi, who translated many Greek scientific and philosophical texts into Arabic, and Thabit Bin Qurrah, who translated ancient Greek philosophers like Apollonius, Archimedes, Euclid and Ptolemy, are among the earliest and the greatest Muslim philosophers.
languages into Arabic, because CA originally was a group of diverse and non-standardized dialects spoken by Bedouin tribes. Since then Arab philosophers began to standardize CA grammar because of the common belief among them that CA was insufficient for the needs of the modern time. They began to systematically create new words, grammatical structures and stylistic elements in order to express new concepts and grammatical structures adapted from foreign languages. In this respect, Abdulaziz points out that

Classical Arabic, which further developed in the Middle Ages, was not equipped to cope with the new conceptual demands. It was now a question of developing within a few decades, in response to the various socio-cultural and psychological stimuli, a new and refined Arabic capable of expressing a material and intellectual civilization that had evolved over centuries in Europe. (16)

In spite of the fact that it is sometimes hard to distinguish between CA and MSA because they share a number of linguistic features, many grammatical rules that are deduced out of CA are still applicable to MSA (Van Mol 2). Nonetheless, they are different in their phonology, morphology, and syntax. MSA is more flexible in its constructions. In other words, grammatical components are distributed differently in the two language forms. For instance, unlike CA, MSA has an acceptable alternative word order—subject-verb-object (SVO; الرئيسي غادر raies khadara “The president left”)—in addition to the unmarked order, verb-subject-object (VSO; غادر الرئيسي kharada a-raies “Left the president”). CA, on the other hand, has a preferred and controlled VSO word order due to cultural-religious insistence on keeping the language of the Qur’an and the tradition pure.

Furthermore, when many Western scholars started to take an interest in Arabic with the establishment of the first Islamic state in 622, their unfamiliarity with CA often resulted in their imposing structures of their own native tongues onto Arabic (Alotaibi 166), such as
passive sentences with expressed agents. The frequent use of these alien structures in informal Arabic media has created the impression, particularly among students, that these structures are acceptable and natural in MSA. Stetkevych states in this regard that MSA deviates strongly from CA. For him, “Modern Arabic is moving away from both the classical and the colloquial languages. While retaining the morphological structure of classical Arabic, syntactically and, above all, stylistically it is coming ever closer to the form and spirit of the large, supra genealogical family of Western culture bearing languages” (121). One of these imposed alien structures is the agentive passive structure. In English, the agent can occur in passive sentences (e.g., ‘The president was assassinated by the leftists’), whereas CA normally uses only the agentless passive. The passive is used when the agent is either unknown or ignored. If speakers want to mention the agent, they must use the active form. “No formal equivalent of the agentive construction is found in Standard Arabic” (Aziz 65). Farghal (142) supports Aziz’s claim, as he points out that there is no natural way in Arabic of mentioning the agent in a passive sentence. Thus, the above English passive sentence could be translated into an active Arabic sentence as follows:

إغتال البساريون الرئيس.

Ekhtal alysariona a-raies

The leftists assassinated the president.

The effect of Western languages, mainly French and English, has resulted in the introduction of new expressions like ‘من طرف’  من طرف، ‘by’ biwaseiat, etc. (the equivalent of the agentive par in French and by in English) to indicate the agent in passive sentences, which are basically agentless in CA. Both agentive and agentless structures are now commonly used in MSA:

أغتال البساريون الرئيس. (active)

Influence of foreign languages, however, is a universal phenomenon. Languages tend to influence one another in different means and for different reasons. For example, English borrowed a great number of words from languages like Latin, Greek, French, German and Arabic.
Ekhtal alysarion a-raies

The leftists assassinated the president. (active)

أُﻏﺘﯿﻞ اﻟﺮﺋﯿﺲ (agentless passive)

Aukhteal a-raies

The president was assassinated. (agentless passive)

أُﻏﺘﯿﻞ اﻟﺮﺋﯿﺲ ﺑﻮاﺳﻄﺔ اﻟﯿﺴﺎرﯾﻮن (agentive passive)

Aukhteal a-raies bewasetat alyسارion

The president was assassinated by the leftists. (agentive passive)

Despite the strong evolution of MSA, in the AW, the term ‘MSA’ is still ill-defined; it has not been widely accepted to label the new Arabic language variety. Many scholars still mix up the two different terms (CA and MSA) and use them interchangeably (see Van Mol 2002). Sometimes it is hard to make a clear distinction between CA and MSA because they are often seen as a unity (القوس al-fosha) (cf. Van Mol 2002, Parkinson 1993, Forkel 1980). Parkinson notes that “Although scholars have no trouble distinguishing MSA from CA on formal grounds, native speakers in Egypt do not typically distinguish between the two, using the term fusha for both” (47). What is clear, however, is that MSA is not the same as CA. CA is the language that has been laid down in the Qur’an, classical literature, and traditional Arabic grammars, whereas MSA is the language of contemporary literature, the modern Arabic media, scientific and technical writing, as well as administration and diplomacy. It is also not accurate to link CA to a certain period6 or to link MSA to a certain level of language use, because CA is still used for many religious purposes and MSA often occurs in dialectal language.7

---

6 Justice (1987: 12) reserves the terms CA for “the period between Muhammad and any of the various eclipses and hard times that succeeded each other beginning around the eleventh century.”

7 MSA is generally taught in schools as a more simplified form (in style and grammar) of CA, where grammar and stylistic structure are more sensitive and more complex. Because of this simplification, the vocabulary of MSA is considered more colloquial, as it involves many words from local dialects.
Furthermore, the fact that MSA is taught in schools almost as a second language means that “the learner [of Arabic] has less reinforcement from the audial pathway. This may represent a significant handicap, as for many people the spoken word imprints better than the written word” (Justice 19). As a result, large segments of Arabic-speaking communities do not have a sufficient command of MSA. The difficulty in obtaining an adequate performance in MSA has led to the appearance of rather significant variations among regional dialects. The similarities and differences between Arabic dialects and their divisions into regional groups as well as how these dialects differ in their incorporation of MSA will be discussed in more detail in the following section.

2.2.2 Arabic Dialects

The Arabic language, like any other language, is shaped by religion, education, politics, social norms and other environmental factors. Yet Arabic has many different regional dialects (e.g., Libyan Arabic spoken in Libya, Algerian Arabic spoken in Algeria, Moroccan Arabic spoken in Morocco, and so forth). These dialects vary to a great extent, but as a unified language, MSA is considered the lingua franca of the entire AW. It is more or less comprehensible for any citizen in these countries with a certain level of education. MSA is, however, structurally and functionally different from the spoken dialects. The current AW can be classified into many different dialectical groups based on morphological and syntactic similarities and differences (see Map 2.1). Bateson (103), for instance, divides the Arabic dialects into two large groups. She distinguishes between the Eastern Arabic dialects, spoken in the Arabian Peninsula and the whole Arabic area east of Egypt, and Western dialects, spoken in the whole area west of Egypt, which includes Libya, Tunisia, Algeria, Morocco, Sudan, Mauritania, and Western Sahara. Her division, according to Van Mol (22), is exclusively based on the variation of one single grammatical element, i.e., the contrast between the prefixes of present tense verbs.
However, a finer distinction can be made based on the variation of not only grammatical elements, but also phonological, morphological and lexical aspects. The contemporary Arabic dialectical world can be divided into five large groups (see Table 2.1):

<table>
<thead>
<tr>
<th>Dialect region</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maghreb Arabic</td>
<td>Morocco, Algeria, Tunisia, and west parts of Libya</td>
</tr>
<tr>
<td>Egyptian Arabic</td>
<td>Egypt, and east parts of Libya</td>
</tr>
<tr>
<td>Levantine Arabic</td>
<td>Syria, Lebanon, Israel/Palestine Authorities, and Jordan</td>
</tr>
<tr>
<td>Gulf Arabic</td>
<td>Kuwait, Qatar, Bahrain, UAE, Saudi Arabia, and Oman</td>
</tr>
</tbody>
</table>

Table 2.1: Regional dialects of Arabic

In spite of the fact that there are many similarities between these dialects (e.g., Becker 161 observes a correspondence of 78% between Lebanese and Egyptian dialects, 65.8% between Moroccan and Lebanese dialects, and 62.2% between Moroccan and Egyptian dialects), it is not possible for people from different dialectical regions to understand each other well, especially those from very distinct geographical areas such as Morocco and the Arabian Gulf. One of the many examples concerning phonological
variation among different Arabic regions is the voiceless uvular stop ﻗﺎف (qaaf), which is used in MSA as part of the language alphabet and phonology in both written and oral forms (see Bahloul 2007). The use of the phoneme qaaf differs from one Arabic dialect to another and has undergone a number of changes within and across various Arabic speech communities. These changes involve a number of phonological production areas of the sound qaaf, such as raising, lowering, and substituting. It has been preserved in a number of dialects such as the Libyan and Tunisian dialects, whereas others, such as Egyptian, make a number of substitutions. The following examples illustrate these variations:

<table>
<thead>
<tr>
<th>English term</th>
<th>MSA term</th>
<th>Libya and Tunisia</th>
<th>Egypt</th>
</tr>
</thead>
<tbody>
<tr>
<td>target disk</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف (alqurs alhadaf)</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف (alaurs alhadaf)</td>
<td></td>
</tr>
<tr>
<td>digital divide</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف الرقمي (al fasil araqami)</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف الرقمي (alfasil arami)</td>
<td></td>
</tr>
<tr>
<td>hard disk</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف الرقمي (alqurs araqami)</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف الرقمي (alaurs arami)</td>
<td></td>
</tr>
<tr>
<td>double click</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف الرقمي (anaqr almizdawij)</td>
<td>ﺑﺒﺴﺮ اﻟﻠﺬف الرقمي (anaar almizdawij)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2: Changes in the phoneme qaaf in regional dialects

The examples in Table 2.2 show the terms target disk (البسبور الهذف), digital divide (البسبور الرقمي), hard disk (البسبور الصلب) and double click (النقر المزدوج) pronounced in different ways: the uvular stop qaaf (ﻗﺎف) is pronounced as a uvular stop [q] in Libya and Tunisia and as a voiceless glottal stop [a] in Egypt. Egyptian dialect replaces the phoneme qaaf in most if not all lexical items, regardless of its phonological environment (see Abdel-Jawad 1981).
The presence of such variations across the majority of Arabic dialects is confirmed by a number of studies. Bahloul, for example, in a survey conducted to examine dialectical variation in conjunction with the geographical distribution of eighteen Arabic dialects from Morocco to Yemen, observes that three North African countries – Morocco, Algeria and Tunisia – exhibit the use of more than one variant of the phoneme qaaf. He notes that “two major variants of the qaaf, namely the uvular [q] and the voiced velar [g], divide urban from non-urban Arabic-speaking communities in Morocco, Algeria, and Tunisia, whereas the lack of such variation in Libya and Western Egypt gives prominence to the velar stop [g] which appears to be a speech feature of all Arabic-speaking communities residing in cities, towns, and/or villages” (272).

To make matters even worse, Arabic has different dialects not only between countries, but also within the same country. In Libya, for instance, there are three major dialect areas:

- Western Libyan (known as Gharbawya, or \textit{Western}), which is spoken in ‘Tripolitanien’, the areas of Musrata, Tripoli and the West Mount. It is similar to the Arabic dialect spoken in southern Tunisia;
- Eastern Libyan (also called Shergawya, or \textit{Eastern}), which is spoken in ‘Kyrenaika’, the areas west and east of Benghazi and extending beyond the borders of Libya into the Western area of Egypt; and
- Southwestern Libyan, which is spoken in Fessan and the whole area south of Tripoli.
Map 2.2: Major Libyan Arabic dialects

It should be mentioned here that besides Arabic speakers, the minority groups in Libya, such as Amazigh, live mainly in the Nefusa Mountains, about 60 miles east of Tripoli, and stretch into the south-eastern area of Tunisia. Other Libyan Amazigh live in the Cyrenaican (Kyrenaika) oases of Augela, Hun, Socena, and Zuara of Tripoltania. They all speak a Tamazight dialect called Zenata or Zanatiyah. In southwest Libya, Tuareg tribes who speak another Tamazight dialect called Tamahak or Senhaja live scattered in the area east of Ghat and around Ghadames. Libyan Arabic and Libyan Amazigh regional dialects are mutually intelligible to most Libyan Arabs and Amazigh respectively, although they vary to some extent. Nevertheless, as a unified language, MSA is considered the lingua franca of the entire country. Most, if not all, Libyan Amazigh are bilingual and speak Arabic as a second tongue.
MSA is, however, structurally and functionally different from these spoken dialects. A number of studies have been conducted to investigate the similarities and differences between MSA and Arabic dialects. The most important difference between MSA and Arabic dialects—and one that all dialects have in common, according to Bateson (97) and Forkel (52)—is the omission of the إعراب إعراب, an Arabic term that refers to the system of nominal and adjectival suffixes used in Arabic. Ibn-Khaldun (1332-1406), one of the most important figures in the field of history and sociology in Muslim history, also reports the omission of إعراب إعراب in Arabic dialects (see Thiry 108). Moreover, due to the influence of these dialects, MSA also shows significant regional differentiation. Diem, for instance, observes that “there are big differences between the MSA of the East and the MSA of North Africa” (2). Farhat (7, cited in Van Mol 2003) shares this view, as he states that MSA is evolving along two major lines, the North African and the Eastern.

This co-existence within a society of two quite different language varieties, each used in different social contexts, provides a prime example of the linguistic phenomenon of diglossia. The following section will deal in more detail with diglossia in the AW in general and the ANAC in particular.
2.2.3 Arabic Diglossia

In an article entitled ‘La diglossie arabe,’ which describes the language situation in Algeria in the 1930s, Marcais (1930) was the first to describe the diglossic status of the Arabic language, although he did not use the term *diglossia* in his description. He distinguished two Arabic varieties occurring side by side: the written language, which had never been spoken and is never spoken, and the spoken language, which had never been written and is never written (cited in Van Mol 41). Later on Ferguson (1959), who is credited with the first use of the term in the famous article ‘Diglossia,’ gave a more comprehensive definition of diglossia by testing it in distinct language situations where the vernacular (which he called *low variety L*) and the formal (which he called *high variety H*) occur side by side—namely, Greece, Arabic-speaking counties, Haiti and Switzerland. He defines diglossia as:

> a relatively stable language situation, in which in addition to the primary dialects of the language (which may include a standard or regional dialects), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation (435).

Unlike Marcais, Ferguson does not make a distinction between written and spoken forms of MSA, but he attempts to accommodate the Arabic language in his definition by defining two varieties: The first is high-variety *H*, used for literary discourse because it is codified and standardized, while the second is low-variety *L*, used for everyday communication and ordinary conversation, which is neither codified nor standardized. He describes high-variety MSA as ‘pure’ Arabic, and low-variety dialects as corrupt Arabic.
forms. Similar to Ferguson’s classification, many Arabic researchers (e.g., Heath (2002), Talmoudi (1981), Ayari (1996), Murtaad (1981)) have addressed this diglossic phenomenon in Arabic. They commonly identify two language spheres for Arabic: the educated world, which is expressed by MSA, and the common spoken world, which is expressed through spoken dialects with no written form. The first (MSA) is prestigious due to its close connection with CA\(^8\), the language of the Qur’an, whereas the second (spoken dialects) is corrupt and seen as inferior to MSA.

![Figure 2.1: Arabic categories (diglossia)](image)

Most, if not all, of these studies, however, have examined the diglossic phenomenon and the dialectal variation of Arabic within a distinct regional dialect, e.g., Heath (2002) on Moroccan dialect, Talmoudi (1981) on Tunisian dialect, Ayoub (1968, cited in Baloul 2007) on Egyptian dialect, Benrabah (2007) on Algerian dialect, and so forth. None of them have examined Arabic dialectal variation across the Arabic World or the main regions. The dearth of investigations on Arabic dialects may be attributed to the fact that Arabic dialects

---

\(^8\) According to Mahmoud (214), “[t]he prestige assigned to MSA stems from its close association with CA eminently represented by the Qur’an.”
have very simplified grammatical rules. Badawi (54, cited in Habash 2012) goes further to claim that “أما العامية فهي عندهم ليست ذات نظام خاص وليس لها قواعد “ (For them [Arab speakers], the dialect has no structure peculiar to it, and it does not have any grammatical rules.)

Figure 2.2: Arabic categories (triglossia)

Although the notion of the diglossic status of Arabic is widely accepted among scholars, Arabic as it is used today in three separate categories (Figure 2.2) is closer to triglossia than diglossia. CA is used for religious purposes and traditional literature and, to some extent, has remained relatively unchanged; MSA is used for education and administration, whilst the spoken dialects are used for all types of speech in everyday purposes. Furthermore, the dichotomic division between high and low variety is too general and too impressionistic (Van Mol 49). It gives only a very general view of the Arabic language situation. Ennaji supports this point, as he states that:

Ferguson’s (1959) classification of Arabic varieties into high and low does not actually correspond to the linguistic situation in Morocco and
the Maghreb at large, for we have three Arabic varieties which are in a triglossic relation (270).

Moreover, Ferguson does not mention foreign languages that dominate the linguistic situation of the AW. The situation in the ANAC can in fact be better described as one of intersection between two developing triglossic areas: one in Morocco, Algeria and Tunisia that involves MSA and French as H languages and regional dialects as L languages; the other in Egypt and Libya, which involves MSA and English as H languages and regional dialects as L languages. Aitsiselmi confirms that “the result of French policies of assimilating Algeria culturally and linguistically have fostered a situation of triglossia, which still exists today, where French, like Standard Arabic, is the H language” (191). Since the linguistic reality in the ANAC region is more complex because of the presence of other languages, the polyglossic situations can be described as follows: one involving MSA, CA, Tamazight, regional dialects and French; the other involving MSA, CA, Tamazight, regional dialects and English. Morsly (254) notes that “Four languages are presently in conflict in Algeria, in the various fields of communication: the so-called Classical or Modern Arabic, dialectical Arabic, Tamazight and French.” For instance, an Amazighan might speak Tamazight at home with her/his family, listen to the news and read books in French or MSA, use CA at the mosque and in Qur’an recitation, and use Arabic local dialects in everyday interactions with Arabs.

2.5 The Role and Status of French

The use of Arabic was well established all over the region until contested by French colonialism. Arabic was the only written language before colonization imposed French, written and spoken, and it assumed an official status. In fact, the presence of French in the region is not a one-time event. It represents a continuing policy that began with the French
conquest of Algeria in 1827. Later, France began to assert its influence on the neighboring countries of Tunisia and Morocco, and in 1884 Tunisia officially became a French protectorate. Immediately after colonizing the whole region, France attempted to change the political and cultural tradition of the people by increasing the spread of French at the cost of MSA and the individual dialects (Talmoudi 22). It also tried to create a territory almost dependent on France by imposing the French language and rejecting everything associated with other languages and cultures. This strategy was best summed up by Gordon, who states that “When the Portuguese colonized, they built churches; when the British colonized, they built trading stations; when the French colonized, they built schools” (7). Algeria offers concrete evidence for this tendency. By 1848 it formally became a part of France, and the French language became the official language. Later, French became the official language of the other two countries in the region (Morocco and Tunisia) and the only language allowed to be taught in schools. Arabic “was completely ousted by the language of the colonizer” (Van Mol 26). It was later, in 1938, declared in Morocco a foreign language by decree (Holt 29).

After achieving independence from France (Algeria in 1962; Morocco and Tunisia in 1965), the three countries undertook a policy of Arabization as an attempt to stamp out the presence of French and restore Arabic and Islamic identity in the region. MSA was immediately imposed as the national language in the constitutions. It became the language of the governments and administration and the vehicle of teaching in national school systems. The policy of arabizing educational systems, for instance, included recruitment of school teachers from other countries such as Egypt, because many Algerian teachers lacked a sufficient command of MSA. According to Madeleine (33, cited in Van Mol 2003), however, this policy in particular had a negative impact on education standards in the long
run due to the differences between Egyptian dialect and dialects of the ANAC region, and this caused communication difficulties between students and teachers.

In spite of the Arabization policy, the French language retains a very strong physical presence in the region. Today, the three countries rank among the ten countries in the world with the most French speakers (État de la Francophonie dans le monde 2004–2006), particularly in the case of Algeria and Morocco, where the status of French is far from that of a mere foreign language. Rabah, an Algerian linguist, states that “the situation of French language in Algeria is unquestionably unique in the world” (cited in Aitsiselmi and Merley 9) because of its widespread use in various domains. In most cases, French is the language of administration, education and a large scope of business and the economy. Despite the fact that these countries gained their independence fifty years ago, French still plays a very significant role in the region. It has continued to occupy an important position because it is seen as the language of social and professional success. “There is a clearly perceived link between speaking French and achieving success: people in power, be it in politics, the army, business, banking, public and private companies, always have a good command of French” (ibid 202). Many business activities, even simple ones, are still conducted using French.

In Morocco, Algeria and Tunisia, French is used together with Arabic as a language of instruction and administration. All official documents such as constitutions, laws and government correspondence are produced in both Arabic and French. Even bills for utilities, taxes, doctors’ reports and prescriptions and most medicine inserts and labelling are bilingual. Only two areas of the administration are completely arabized, namely, the justice system and the Ministry of Islamic Affairs.

The influence of French can also be observed in the education systems. In spite of all efforts to arabize higher education since independence from France, the university sector
has remained far from fully arabized. In the humanities, however, arabization has progressed further than in science subjects; many scientific fields of knowledge in universities and colleges, particularly in Algeria and Morocco, are still taught in French. These include, but are not limited to, medicine, pharmacy, engineering, nursing, physics, chemistry, biology, mathematics, and administrative sciences. French is also used, in many instances, for everyday activities. As a result, Arabic dialects in the region have incorporated so many terms of French origin in order to make up for scientific and communicative lacunae in their vocabulary that one needs dictionaries to list them all.⁹

Moreover, dramatic improvements in technology and communications have enabled greater exposure to the French language. People in the region are increasingly exposed to French language and culture through movies, satellite channels, recorded materials, the Internet, etc., which means that everybody is regularly exposed to French language and culture. Therefore, “people in North Africa are aware, as never before, of the language, culture and civilization of France” (ibid 186). The French language also features regularly in a number of types of broadcasts. French films, documentaries and series are shown frequently, without dubbing or sub-titling.

Not surprisingly, French currently holds the leading position among the languages that have influenced Arabic in the ANAC region. This influence can be clearly seen in people’s dialects, which are often mixed in with French in everyday conversation to form a dialect often informally called Frarabic. Frarabic is “An amalgam of the French and Arabic languages, sometimes producing very distorted sounds. It relates mainly to code-switching in mid-sentence or a single French word that has been given Arabic intonation” (Urban Dictionary). Frarabic is often used in Francophone Arab countries such as Morocco, Tunisia

---

⁹ For example: The Dictionary of Mixed Dialects and Foreign Words in Arabic by Philip Brown (2001) and Dictionary of Medical Terms: English - Arabic, with a French and English glossary of terms with different roots by Ibrahim and Mansour 2006.
and Algeria. In Tunisia, for instance, the two greetings of Arabic السلام عليكم (equivalent to hello in English) and French ça va are often mixed as Asslama, ca va. Also, شكراً shukran and merci (thank you in Arabic and French) are used interchangeably. Similarly in Algeria, according to Aitsisiselmi and Marley, the use of mixed code is more marked especially when the subject is culture-specific; “there is frequent use of Arabic terms which signal a religious reality for Algerian listeners that cannot be adequately expressed in French” (204). For example, the Arabic words ربي rebbi my God and الحمد الله il-hamdullah thank God are often used in expressions such as on réussira b-rebbi (we’ll succeed, God willing) and il se porte bien maintenant, il-hamdoullah (he is feeling better now, thank God) (Aitsiselmi and Marley 204).

To conclude this brief review about the status and role of French in the ANAC, it is obvious and understandable that the Arabization policy adopted by the ANAC countries since independence did not bring an end to the use of French. French still has an important linguistic and cultural influence in many significant domains, especially in administration, education and the media, although more than a half century has passed since the end of the colonial era. Although people have more exposure than ever before to the Arabic language through education and media, French retains a very strong linguistic and cultural presence due to the heavy investment in the domains of media and education on the part of France. French authorities remain aware of the value of education and the media as means of maintaining the francophone influence in the region. They do not appear to be aware, however, that this status cannot be controlled; the improvement in media, telecommunications and the rapid growth of the internet has introduced English as a third powerful competitor in the linguistic landscape of the region. The emergence of English as an alternative option that gives access to advanced technology and science, as well as the shift in
attitudes in favour of English, particularly on the other side of the region (Libya and Egypt), will be discussed in the following section.

2.6 The Role and the Status of English

The use of English, especially in the western region of the ANAC, has always played a secondary role due to the heavy dependency on the French language in spoken and written discourse. This may be explained by several factors. First, the official attitude toward English, although unstated, is something like the following: ‘It’s their (U.K. and U.S.) language; if they want to promote it – like the French promote their language – they will help us financially.’ In fact, France has always been reluctant to relinquish its linguistic influence in the region, because it realizes that loss of language is accompanied by a decline in political, economic, and cultural ties (Battenburg 285). Second, there is a continuing resistance to the spread of English in an economic and educational system that is dominated by the French language in Morocco, Algeria and Tunisia. In these countries, French is favoured because it connects the national economy to a great part of the international markets. Economic and educational activities with other non-English-speaking countries in the European Union (France, Italy, Germany, etc.) are given special importance by governments and businesses in this region. English-speaking countries such as the United Kingdom and the United States are of secondary importance.

The situation in Egypt and Libya is quite different from in Tunisia, Algeria and Morocco. Although Egypt was conquered by the French from 1798 to 1801 and French was the first major European language introduced to Egyptians during Napoleon Bonaparte’s campaign, French has had little impact on the dominance of Arabic among Egyptians, although it was favoured as a foreign language among other choices such as Turkish, Persian and Italian. This foreign language situation changed after the British occupation began in the
1880s, particularly at the beginning of the First World War. The British made concerted efforts at widening the influence of English instruction in Egypt’s schools because, as Tignor (326) puts it, they saw Arabic as ‘too imprecise’ to be a language of science. Therefore, all secondary school and university subjects, except for Arabic and mathematics, were taught in English. This policy continued until the British administrator, Lord Cromer, reversed the stance of the education bureaucracy and the repression of Arabic in schools in 1905 (Schaub 227).

In Egypt today, English is a mandatory subject that is first introduced at the preparatory level, and it is the language of instruction in most faculties in national universities. Massialas and Jarrar (1988) observe that many of the textbooks and reference materials are in English, and in medicine and engineering, all of the texts and many of the lectures are in English. English also plays a significant role in the popular culture of Egypt; it is used alongside Arabic in road signs and the names of shops and cafés. Printed materials in places such as hotels, restaurants, banks, airports, travel agencies and post offices are usually written both in English and Arabic. Brand names, such as KFC, Subway, and Starbucks, are written only in English and without any kind of translation. In fact, the tourism industry, which is one of the most important sectors and the largest source of hard-currency revenue in Egypt, may be considered the main factor behind the dramatic widespread study and use of English. Elkhatib (1984) notes that Egyptians, especially if they live in areas frequented by tourists, need to learn English. Merchants and even juice sellers stand to increase profits greatly if they are able to communicate with visitors in English (cited in Schaub 229).

Although the use of English in Egypt is increasingly becoming common at both formal and informal levels, Arabic – Egyptian Arabic in particular – is without question the dominant language. It is the language of choice for most Egyptians, not only when they speak
in their everyday communication, but also in more formal contexts such as communication in higher education, business and policy.

Since Libya’s independence in 1951 and until the mid-1980s, the English language enjoyed a significant status in that country. Italian is well understood, especially by older generations, because Libya was subject to Italian colonialism (1911-1943) and of course there are linguistic vestiges from that period, but today English is the main foreign language. After independence in 1951, Libyan authorities realized the need for Libyans who are capable of speaking and understanding English as a common means of communication. Thus, the overall policy of the Libyan Ministry of Education was to provide English training as a compulsory subject from elementary level through the preparatory and secondary levels to tertiary level in all institutions. However, the plan to introduce English into the elementary curriculum was abandoned in the 1960s due to the critical shortage of qualified teachers (for further details see in particular UNESCO, 1968).

The discovery of oil in the late 1950s also played a major role in the development of the English language in Libya. Major foreign petroleum companies such as Shell, British Petroleum (BP), and Exxon Mobile rushed to Libya for exploration and drilling. By the end of 1977, there were forty-two foreign companies conducting exploratory and drilling activities in Libya (Hassan 2009 http://sepmstrata.org/Libya-Hassan/Petroleum-History-Libya.html). Therefore, the oil and gas industry has increasingly become the main Libyan job market where English is the means of communication between Libyans and native and non-native speakers of English alike. Upon entering the industry, all workers are required to have at least a basic knowledge of English for effective and safe operations (National Oil Corporation 2009, http://en.noclibya.com.ly/). This, in turn, brings a very significant requirement for English language learning and training, particularly with regard to technical
operations. English is, therefore, seen not only as the key to securing a better job in the Libyan oil sector, but also as the means of developing social, economic, commercial and scientific relations with non-Arabic companies and individuals from within Libya and internationally.

However, the status of English has undergone tremendous changes in the past decades for purely political reasons. Influenced by Gamil Abdul Nasser’s revolution in Egypt in 1952, Libya proclaimed a Cultural Revolution in 1973. There was a move to shift authority from Western-oriented capitalism – also called anti-Western – into a strongly nationalist and socialist country (which was at the time perceived as ‘anti-Western’). Thus, everything originating from the West and from the United States and the UK in particular became unacceptable and prohibited, including language, which was regarded as imported culture that had to be rejected. As a result, English books, magazines and newspapers and even Western musical instruments were collected and burned in public squares. These measures were extended, according to Maghur (5), to private schools and foreign centres; several private foreign schools and foreign centres such as the British Council and the American Cultural Institute were shut down. As a result an acute problem now exists, whereby Libyan schools and universities in general do not have a satisfactory stock of books and other references related to foreign language teaching and learning. In fact, most, if not all, Libyan students depend on their teachers to provide them with the books and references they need. They always express dissatisfaction and complain about the lack of sources.

Since the early 1990s the status of English in Libya has deteriorated considerably as a result of political tension between Libya and the West. The USA and UK air raid on Libya on 15 March 1986 and the accusation against the Libyan regime in connection with the bombing of a Pan Am flight over Lockerbie in Scotland on 21 December 1988 led to Libya’s political
and economic isolation for almost a decade. As retaliation on the part of Libyan authorities, in 1986 the teaching of English was banned in schools and universities across the country, and English was replaced by Russian as the primary foreign language in education. Nevertheless, teaching in certain science faculties, such as medicine, pharmacy and engineering, remained mainly in English (Maghur 5). Libyan English teachers were also ordered to teach other subjects such as history and geography. Non-Libyan teachers and professors of English and even French, especially those from Western countries, were notified that their contracts would not be renewed for the following school year. After the government realized the error of this decision and decided to re-incorporate English into the curriculum in 1997, non-Libyan instructors were replaced by Libyan graduates who were neither qualified nor well prepared to carry out the task of teaching English.

The negative consequences of this isolation with regard to the English language became evident over time. The policy led to inadequate standards of English teaching and a lack of trained English teachers, a phenomenon which lasted for the next few years. Orafi and Borg state that “English language teachers in Libya typically graduate from university with undeveloped spoken communication skills in English” (Orafi and Borg 251). The same view is found in Alhmali, who points out that “a common feature shared by the majority of graduates from the English departments of Libyan universities is their undeveloped listening and speaking skills” (20).

Today, the English language is booming in Libya after the National Transitional Council of Libya (NTC) restored Libya’s relations with the West in 2012. Following the government decision to reintegrate English into the curriculum of Libyan schools in 1994, many decisions have been made to improve the quality of English language teaching and learning. These include the decision to start teaching English in Libyan schools from fifth
grade instead of seventh grade and the decision to send students abroad, especially to the
United States and the United Kingdom, on government scholarships. Nevertheless, many of
these students were unable to obtain admission, mainly due to their lack of English. Thanks
to the vast improvement in mass communications, however, the Libyan people, particularly
the younger generations, are more exposed than ever before to the English language through
satellite television and the Internet, as well as to electronic and printed materials (videos,
newspapers, magazines, etc.), and thus they are aware, as never before, of Western culture
and civilization.

2.7 Indigenous Languages

In addition to Arabic, French and English, the indigenous languages of the ANAC
(mainly Tamazight) are also gaining a higher profile in the region. They are found in
Morocco, Algeria, Tunisia, Libya and the Western part of Egypt. The exact population of
Amazigh speakers is hard to ascertain because ANAC countries do not record Amazigh
population figures in their census data. Government population statistics usually group Arabs
and Amazigh together. Chaker (1995), however, reports that the number of Amazighes in the
region ranges from eight to thirteen million. The share of Amazigh population in the ANAC
is 18-25 percent in Algeria and about 40 percent in Morocco, about 1 percent in Tunisia, and
10-20 percent in Libya (see Kratochwil 39). No statistics can be found for Egypt and
Mauritania. In this regard it is worth mentioning that there has been a considerable debate as
to whether Amazigh language is a single language or a group of languages such as Tachelhit,
Tamazight and Tarifit (see Elmadlaoui 2003, Brenzinger 2007 and Sadiqi 1998). According
to Brenzinger,

In order to foster a common identity and join their strength, Amazigh
activists generally speak of one [T]Amazigh[t] language by
acknowledging the existence of regional varieties. While most linguists agree on a close genetic relationship, they distinguish various [T]Amazigh[t] languages based on linguistic analysis. Only neighbouring Amazigh speech varieties are mutually intelligible and linguistic dissimilarity increases with geographical distance (124).

Tamazight, however, is generally accepted as the indigenous language of the ANAC before the arrival of the Arabs from the Arabian Peninsula in the 7th century and the later arrival of people from Europe.

The term “Berber” is actually of Greek origin (barbaros) and was used by the Greeks to designate foreign peoples speaking languages other than classical Greek (i.e., foreign languages to the Greeks). Later on, this term was used by the Romans to designate the peoples in North Africa who did not speak Latin. Many Amazighes reject this term because of its negative association with barbarian. They prefer the name Amazigh, which literally means ‘free man.’ Nowadays, the word Berber is used by foreigners to refer to the inhabitants of the areas where the Amazighes’ languages, mainly Tamazight, are spoken, and the term Amazigh is currently gaining acceptance. According to Aitsiselmi and Marley (192), “This term [Amazigh] is used in the modern Berber cultural movement which attempts to unite the different tribes and varieties, in their struggle for cultural recognition and survival.” It is for these reasons that I choose to use the terms ‘Amazigh’ to refer to Berbers and ‘Tamazight’ to refer to all Berber languages in this study, as noted above (see footnote 3).

The Amazigh live not only in the ANAC but across a much wider area of North-West Africa, extending from a remote isolated group in Mauritania, the Zenaga, to Siwa Oasis in the east of Egypt and down into the Sahara. Map 2.4 shows the distribution of the Amazigh today across North-West Africa.
Although Tamazight is now found on the Internet and used in radio broadcasts, magazines, and newspapers, it is not acknowledged by institutions in any ANAC countries. In Libya, for instance, its usage has been restricted to family circles until recently. Although there is a graphic representation for Amazigh known as *tifnag*, it is going through a static phase due to a lack of use as a written form over the centuries. This has contributed to the development of another form of diglossic situation, as it means that MSA is obviously the language of education and administration, and thus the high-variety language, whereas Tamazight is related to the low domains. It is noteworthy, however, that in Morocco, where the majority of Amazigh (about 5.7 million) are found, Tamazight usage is more extended. In addition to its use in everyday life, it is represented in some public administration when the staff include Amazigh.

In Libya, the strong wave of nationalism accompanying the coup of September 1969 found expression in a campaign designed to elevate the status of Arabic. As a result, the Tamazight language has been officially excluded from prestigious linguistic domains for
many decades by Libyan politicians in favour of Standard Arabic. The former Libyan leader Mu’ammar al-Qadhafi was always hostile towards the Amazigh and their language. For him, “the will to use and maintain the Berber [Tamazight language] is reactionary, inspired by colonialism ….” He neglected any distinct Amazigh identity and considered them as being عرب اقحاح arab aqhah pure Arabs. In a speech to Tuareg tribal leaders in Niger in 2007 al-Qaddafi went so far as to deny the existence of Amazigh in the whole of North Africa. He stated that “Berbers are the Arabs that came via البر barr [land in English] …. then colonialism arrived and said that Berbers are a different people from the Arabs.” Although there is a significant large population of Amazigh in Libya, the 1977 Declaration of the Establishment of the Authority of People (DEAP) repeatedly places emphasis on the Arab nature of Libya. Even the 11 December 1969 Libyan constitutional declaration (replaced later by the DEAP) defines Libya as an Arab nation and stresses that Arabic is the country’s only official language. This requires that all street signs, shop window notices, signboards, and traffic tickets be written in Arabic. There is no mention of the Tamazight language, and accordingly Tamazight was not recognized, despite the fact that it was, and still is, a living reality in Libya.

The existence of the Amazigh and their identity is also ignored in Egypt, Tunisia and Mauritania in favour of the Arabs and because they are few in number in these countries. Morocco, in spite of its small geographical size (about 447.000 square kilometres), has the highest Amazigh population (about 5.7 million) and the greatest variety of Tamazight dialects (Tashelhit in the south, Tamazight proper in the central area, and Tarifit in the north). Brenzinger points out that Amazigh from the southern part of the Tashelhit area can communicate with those speaking the central Tamazight variant. The latter can also converse with the speakers of Tarifit in the north, but Tashelhit and Tarifit dialects are mutually unintelligible. Although Morocco has the most linguistic diversity and the linguistic situation
there is often described as complex, it seems that the country has the most positive prospects for the future of Tamazight, not only because the Amazigh population represents an important part of present-day Morocco, but also because their languages and culture are often encouraged by Moroccan authorities. The Institut Royal de la Culture Amazighe (The Royal Institute of Amazigh Culture, or IRCAM) was founded in the capital Rabat by King Mohammed VI on October 17, 2001 to preserve and promote Tamazight languages and culture. The goal of IRCAM is to standardize the Moroccan Aamazigh dialects and to establish a standard Tamazight orthography, which in turn “will allow for the introduction of the Tamazight language in the educational system and ensure the spread of its influence in the social, cultural, media, national, regional and local contexts” (IRCAM 2004, http://www.ircam.ma/). This tolerance towards Tamazight language lies in the fact that the Moroccan Amazigh constitute a significant part of Moroccan’s population, about 40% of the population (see Table 2.3).

2.8 Language Competition and Language Attitude

As noted above, the ANAC has long been mistakenly characterized by the omnipresence of Arabic, but it is considered today to be perhaps the most multilingual region in the AW. The situation is best characterized as multilingualism in the sense that many languages and varieties are used in different domains: specifically, MSA, colloquial Arabic dialects, French, Tamazight, and recently English. Although the entire region is Arabic-speaking and MSA is the only official language, the current sociolinguistic situation is vastly more complex and more dynamic than in other Arabic regions such as the Arabian Gulf and the Middle East, simply due to the linguistic diversity of the region. An example of such linguistic diversity is Morocco, which hosts several different languages in spite of its small size. The linguistic situation in Morocco is often described as complex, perhaps for good
reasons. For a country with an estimated population of over 29 million and with over 6 languages, Morocco is not only the country that has the highest population density in the ANAC, but also one of the countries with the highest linguistic diversity. It is largely a multilingual country with languages of unequal social, official and educational statuses. Although Arabic is the only recognized official language in Morocco, French is commonly used on many formal occasions due to France’s colonial occupation of the country for nearly 150 years. Moroccan Arabic is spoken natively by 19 million people, i.e., 65% of the population, and a further 20% speak it as a second language (Gadelii 15).

In general, the majority of the current population of the ANAC is largely composed of Arabs and Amazigh. However, there are significant European populations in the region, such as Spanish and French. Table 2.3 gives an overview of the linguistic situation across the region:

<table>
<thead>
<tr>
<th>Country</th>
<th>Population</th>
<th>Area in sq. km</th>
<th>Linguistic situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morocco</td>
<td>29 million</td>
<td>447,000 km²</td>
<td>19 million Arabs (65%), 5.7 million Amazigh (40%), 80,000 French native speakers, 20,000 Spanish native speakers</td>
</tr>
<tr>
<td>Libya</td>
<td>5.7 million</td>
<td>1,800,000 km²</td>
<td>80% Arabs, 17% Amazigh, others (including Greeks, Maltese, Italians, Egyptians, Pakistani, Turks, Indians, and Tunisians) 3%</td>
</tr>
<tr>
<td>Algeria</td>
<td>35.5 million</td>
<td>2,400,000 km²</td>
<td>72% Arabs, 27% Amazigh, 1% Europeans (mainly French)</td>
</tr>
<tr>
<td>Tunisia</td>
<td>9.6 million</td>
<td>163.6 km²</td>
<td>98% Arabs, 2% Amazigh and Europeans.</td>
</tr>
<tr>
<td>Egypt</td>
<td>82.6 million</td>
<td>1,001,450 km²</td>
<td>99.6% Arabs, 0.4% other</td>
</tr>
</tbody>
</table>

Table 2.3: Distribution of languages by country

One of the many factors that have made developing an appropriate Arabization policy harder to achieve is the ongoing competition and conflict among Arabic, French, English, and Amazigh. The widespread official realization that Arabic is the only politically,
economically, and socially acceptable solution and the increased prominence of French and
English as a medium for modern science and technology and for economic development, as
well as continued calls to use the Tamazight language and its widespread use in radio
broadcasts, magazines and newspapers, have all resulted in competition and conflict among
these languages for prestige and power. This competition for prestige, especially among the
three colonial languages (Arabic, French, and English), is an increasingly noted phenomenon
and in turn has resulted in a linguistic taxonomy regarding the relationship among these
languages and their political status. As far as the current relationship among the languages in
the region and their social functions is concerned, the linguistic situation can be classified
into two functional groups: (1) institutional and administrative languages, which include
those languages with a mainly formal usage (high variety H), i.e., MSA, English and French;
and (2) communication languages, which include commonly used mother tongues (low
variety L), i.e. Tamazight, Libyan Arabic, Moroccan Arabic, Tunisian Arabic, Algerian
Arabic, and Frarabic (a mixture of French and Arabic dialects).

Tension exists not only between Arabic and European languages represented by French
and English, but also within the ANAC context between Tamazight and Arabic languages
and cultures. Tamazight has been and often still is regarded as a threat to Arab Muslim
identity (Boukous 24). For this reason, according to a report about the Amazigh of Libya that
was submitted in 2004 to the Committee for the Elimination of Racial Discrimination
(CERD), Arabic (the language of the Qur’an) had succeeded in weakening Tamazight and
seriously threatening its survival, although it has been able to resist almost all the conquering
empires (Phoenicians, Byzantines, Romans, etc.). Ironically, it is not MSA that threatens the
ethnic tongue of Amazigh. Because Arabic dialects have become the mother tongue of the
Amazigh of the ANAC, the main threat to Amazigh is colloquial dialects, not MSA. The
negative attitude toward Amazigh held by officials is also reflected in the general public’s
view. For most Libyans and the ANAC people at large, Tamazight is a dialect that is neither a language of wider communication nor a language with a rich written literature. These are seen as grounds for believing that it is not worth introducing Tamazight into schools because it is neither standardized nor codified.

However, Tamazight, which has been marginalized in Libya for centuries, has in recent years experienced a dramatic change in fortune thanks to the current economic and political reforms and the emergence of democratic trends. The geopolitical re-composition of the AW in general and of the ANAC in particular (in Tunisia, Egypt and later in Libya) and the emergence of new democratic trends will possibly make room for the reassertion of a wide range of local and regional identities that had been suppressed for decades and that give rise to new linguistic contexts. In Libya today, for instance, the Tamazight language, which has been officially excluded for many decades by Libyan politicians, has a historical opportunity to be recognized as an official language in the new Libyan constitution, despite the fact that the NTC made no mention of Tamazight as an official language in its recent Constitutional Declaration of August 2011. After the 17th February uprising against the dictatorial Gaddafi regime, the Tamazight language has in fact gained significant status with Arabic in Libyan’s media because of the Amazighs’ pivotal role in the uprising. In a short time, a number of Amazigh printed and online journals and publications have sprung up in Libya and abroad, with radio and TV stations broadcasting in both Arabic and Tamazight. This new trend is summed up by Peter Graff, a correspondent for Reuters News Agency, who was reporting from the Amazigh town of Jadu during the uprising:

For a few weeks, a radio station has been broadcasting from here in both Arabic and Amazigh, in what Berber activists believe are the first conversations in their language over Libyan airwaves in four decades. An
Amazigh publishing house has printed four books so far over the past month, billed as Libya’s first publications in the language since Gaddafi seized power. (http://news.nationalpost.com/2011/07/11/ancient-language-renewed-in-libyan-rebellion/)

Many argue (Salmi, 1987; Zizi, 1984, Bentahila 1983, Ennaji 1988) that the divergent nature of Arabic and the abovementioned state of multilingualism are not the only sources of the problems in developing an appropriate Arabization policy. They argue that the attitude towards specific languages is also to blame. Ennaji (1988) suggests that many problems that the Arabization process in Morocco is encountering can be attributed to contradictory attitudes held among policy makers and ordinary people. The attitude toward the role of languages is rather ambiguous. Whilst for some people French is essential for economic and social development, for others, re-establishing Arabic as the sole national and official language should be the priority of language policies. The first conviction was held by technocratic modernists, who believe that bilingualism must be maintained in order to retain access to the Western world and technical material, which would not otherwise be possible. The second belief was held by traditionalists and Islamists, who consider French to be a remnant of imperialism and one of the major reasons behind the denationalization of the countries. Benrabah identifies a third tendency, ‘multilingualism,’ which calls for the recognition and promotion on an equal footing of all languages used, i.e., French, Arabic and Tamazight (347).

In fact, French has long been associated with modernity, especially in economics and technology—domains in which English is introducing competition—whereas Standard Arabic is associated with ‘Arabo-Muslim tradition’ (Mufwene 2008). French has been viewed as the key to improving one’s social and economic standing, and obviously it
constitutes a greater asset on the local job market than English and Arabic, especially in jobs that are more highly valued, such as in the fields of economics, engineering, medicine, and higher education. Aitsiselmi and Marley (202) observe that “there is a clearly perceived link between speaking French and achieving success: people in power, be it in politics, the army, business, banking, public and private companies, always have a good command of French.” Studies (see Benrabah 2007, Marley 2000, Bentahila 1983, and Mouhssine 1995) indicate that most people, particularly the younger generations who have grown up in an arabized education system, do not value monolinguals and the mastery of Arabic over other languages. Most of them see French as an advantage in everyday life and want to maintain it as a tool for communication. In a recent survey, Benrabah found that when the respondents were offered English as an alternative to French, they “seem to reject policies that seek to replace French in favour of English.” (245)

However, the growing demand for English as a tool and means for modern science and technology and for economic development, as well as the efforts of some English language trainers such as the British Council and AMIDEAST, have contributed to a shift in the linguistic orientation of many francophone regions, and the ANAC is not an exception. While English is gradually gaining prominence in many of these countries, French is increasingly losing its political, economic, and social dominance (Battenberg 1). Recent studies (Elbiad 1985, Sadiqi 1991, and Bourenane 1984) have found that English is preferred over French as a foreign language, thus the countries of the ANAC have sought to decrease their dependency on the French language, particularly in government ministries and educational institutions, in favour of English.

Nevertheless, one may say that French will continue to play an important political and economic role in the future despite the increased perception of English as a potential
linguistic option. One obvious example of this is the current heavy dependence of economic and educational activities in the region on French and the significant presence of immigrants from Algeria, Morocco, and Tunisia who travel back and forth between France and the region. Today, for instance, there are approximately four million from the region who live in France, making them a strong minority at approximately 8% (Tarwater and Salih 9).

2.9 Summary and Conclusion

This chapter has described the evolution and current status of the sociolinguistic landscape of the ANAC. The chapter provided a brief review of the history of Arabic language and its development from classical form into modern form. It also examined the status and role of French, focusing on how French, despite having no official status, is far from being a mere “foreign language,” especially in Morocco and Algeria, where French is widely perceived as a natural part of the linguistic landscape and regarded as a language of economic and social promotion that gives access to advanced science and technology. The chapter also indicated the emergence of English as another linguistic option for modern science and technology, despite the privileged status that French enjoys. Finally, the chapter provided insights into changes in language attitudes and use in the ANAC.

From what preceded, it is possible to divide the linguistic map of the ANAC region into Anglophone and Francophone areas with regard to foreign language dominance, and thus it can be divided into two different educational orientations: institutions in Egypt and Libya eager to experiment with English language instruction, and others in Morocco, Algeria, Tunisia, and Mauritania, which have sought to intensify educational instruction in French. This competition between English and French has created two terminological sources: a French orientation in Morocco, Algeria, Tunisia, and Mauritania and an English orientation in Libya, Egypt, and almost all of the rest of the AW. This means that there are two different
language sources for terminology, and the pace and scope of Arabisation has depended largely on which of these languages is more dominant and has more power than the other. In any event, any term is bound to be influenced by one of the source languages, i.e., French, English or local dialects. Consequently, terminological variation may occur as a result of borrowing and loan words from these different language sources, posing a potential communication problem for speakers from different parts of the region, as they may wrongly assume different meanings for each alternative term. Lakhdar-Ghazal admits that “many problems are encountered in translation, chiefly the problem of terminology, which differs from one Arab author to another and from one Arab country to another” (137, cited in Ennaji).

Chapter three will provide a survey of conceptual and terminological definitions in the fields of LP and TP and the relationship among them. A number of critical approaches to TP (such as top-down and bottom-up) will be reviewed. The chapter concludes by reviewing the history and current status of LP and TP in the ANAC. The discussion involves two critical periods of Arabization policies, i.e., LP and TP in the period of colonization and after independence (post-colonization).
Chapter Three

Language and Terminology Policies: Theoretical and Comparative Background
3.1 Introduction

Considering the multilingual situation of the ANAC, which has been discussed in detail in Chapter Two, this chapter attempts first to provide terminological definitions in the field of language planning (LP) and terminology policy (TP), which today have become complementary activities, and to discuss the key stakeholders and elements involved in these two related fields. The chapter also reviews some critical approaches and contributions in LP and explores the relationship between LP and TP, with the aim of locating TP in a LP framework. However, since the focus of this study is on terminology planning, special emphasis is placed on examining the literature on terminology planning. Only a short discussion of the theory of LP will be presented. The chapter also investigates the applicability of different strategies put forward by Arabic scholars in order to deal with the flood of new terminologies being created almost every day. The focus will be on the main methods of creating new terms (i.e., derivation, compounding, loan translation, Arabisation, abbreviation). Each method will be defined and examined separately, and examples will be given. The chapter concludes with a broad review of the terminological situation in the ANAC region and attempts to demonstrate how terminology work is responding to specific challenges raised by the transfer of modern scientific and technical knowledge, particularly with regard to computers and the Internet.

3.2 Language Planning

LP is an old phenomenon, although it is only relatively recently that it has developed as a separate academic discipline (see e.g. Fishman, 1968; Fox, 1975; Rubin & Jernudd, 1971). “[LP] was an integral part of nation building and, in the eighteenth and nineteenth centuries, intellectuals in the United States, France, Germany, Italy and to a lesser extent Britain produced a rich literature on the subject” (Wright 2). However, scholars are not consistent in their use of
terminology when they refer to this field. Many terms have been used to refer basically to the same kind of activities attempting to solve language problems on a national scale. Springer (54), for instance, uses the term ‘language engineering’\(^\text{10}\) to refer to the efforts to standardize the semi-standardized language in the Soviet Union. Noss (1967) uses ‘language development’ in examining the language situation in Southeast Asia. The term ‘language planning’ itself was first launched in 1959 by Haugen in his discussion of the language situation in Norway. Haugen defines LP as “normative work of language academies and committees, all forms of what is commonly known as language cultivation and all proposals for language reform or standardization” (287). He also uses the term to mean “the activity of preparing a normative orthography, grammar, and dictionary for the guidance of writers and speakers in a non-homogeneous speech community” (8). Influenced by Haugen’s definition, Fishman (79) uses the term LP to refer to “organized pursuit of solutions to language problems, typically at the national level.”

In the Arabic literature, there has been no clear agreement among linguists on what the concept of LP exactly means, so Arabic linguistic and terminological organizations use terms such as تخطيط اللغة takgīf al-lugha language planning, سياسة اللغة siasat al-lugha language policy, إدارة اللغة idarat al-lugha language management, تطوير اللغة tāweer al-lugha language development, etc., interchangeably. In addition, there are many related terms, such as ممارسة اللغة language practice—the ecology of language, which, regardless of policies and beliefs, focuses on the actual language practices in a community, on “all the varieties that make up the sociolinguistic repertoire of a community” (Spolsky, cited in Kiss 277)—and إيدولوجية اللغة idīlḥiat al-lugha language ideology or belief, which refers to “the human

\(^{10}\) The term “language engineering” becomes problematic in the 1980s because it was applied to the use of computer technologies to study and control language use, as opposed to efforts to modify language itself (Sue Ellen Wright, personal communication, February 2, 2013).
understanding of language in the form of ideas, beliefs, consciousness about the relation of 
language, and social life” (Gal 319).

From this rich literature on the topic, I have selected Grin’s (18) definition as the most 
applicable to this study. Adapted from Cooper (1989: Chap. 2), Grin defines LP as “a systematic, 
rational, theory-based effort at the societal level to modify the linguistic environment with a view 
to increasing aggregate welfare. It is typically conducted by official bodies or their surrogates 
and aimed at part or all of the population living under their jurisdiction” (the italics are mine). 
This definition implies that LP is assumed to be consciously and systematically organized in 
order to achieve specified objectives in a given ‘linguistic environment.’ The objectives include 
protecting a minority language and increasing its visibility in cultural life, developing foreign 
language skills in a certain population, or even suppressing the use of certain languages, welfare 
in Grin terms (28). LP is often the plaything of language forces (Fishman 383). Language forces 
such as ‘state authorities and their surrogates’ (‘official bodies’ in Grin’s terms), but anyone 
from individuals to multinational companies, can engage in or support LP. However, LP often 
requires political decision and official deliberate intervention in order to be effective and to 
achieve ‘welfare’ because, as Grin (29) puts it, in the absence of deliberate state intervention, 
initiatives on the part of individuals and even the private sector would give rise to behaviors that 
would not result in the linguistic environment desired.

Planning is necessary for many languages to strengthen and develop their status and thus 
allow them to function effectively either domestically or internationally. Drame emphasizes that 
“In today’s global information society and increased interaction, language has to be formed, 
planned and standardized otherwise it will never become an effective medium of communication 
for a larger society. Without deliberate efforts it will become extinct soon because it will be
dominated by another” (229). I argue that this statement holds true for languages other than English and, to a certain extent, French. Although English is not an official language in every country in the world, nor is it spoken by people everywhere, it has successfully achieved the status of a global medium of technical and scientific communication without formally recognized LP or policy. Other languages (e.g., Hungarian and Japanese) have implemented rapid planning to meet the demands of globalization and modernization. In Japan, for instance, the modernization and simplification of the major Japanese scripts, as well as the establishment of guidelines on the use of these scripts, have been planned and implemented, to a large extent, through the Japanese government’s LP (Gottlieb 27). Some other languages (e.g., Arabic and Hindi) have failed to exhibit any significant lexical expansion. Obviously, a language that is limited to everyday vocabulary and cannot meet the requirements of modern science and technology remains a language of the home and runs the risk that it will never be, or will cease to be, a language for science and industry (Ibarahim 2). Therefore, a systematic TP based on consolidated LP is urgently needed for the Arabic language in order to achieve a significant level of development and modernization.

3.3 Language Policy vs. Language Planning

Since the early 1980s, the interrelated concepts of language policy and LP (LPP) have been a major focus of research in the general language disciplines, particularly in sociolinguistics, politics and applied linguistics. Cooper (31) considers language policy to be comprised of political decisions or orientations to guide and implement LP proposals. He defines it as “a process of decision-making concerning the teaching and use of language and the careful formulation by those empowered to do so, for the guidance of others.” Spolsky distinguishes two other related components besides LP: language practices and language beliefs. The former
concerns the choices of community members among varieties available for use, whereas the latter includes attitudes and beliefs towards these varieties.

### 3.4 Language Planning Typology

Heinz Kloss (1952) is credited with distinguishing between *corpus planning* and *status planning*. For him, *corpus planning* refers to efforts to change the lexicon, grammar, phonology, orthography and writing system of a given language, whereas *status planning* is the planned efforts to change the societal function of a language at governmental, educational, mass media, legal and religious levels. Ferguson (21) makes the same point; for him, corpus planning is intended to change the code or the form of the selected language. Status planning, on the other hand, is intended to address the functions of language(s) in society. Similarly, Ager (10) makes the same distinction between status and corpus planning.

Although status planning has attracted greater attention and the interest of both political scientists and sociologists, since it seeks to extend the domain of a given language to new functions and also to enhance its prestige and ensure both the ascendancy and primacy of the language (Parker 181), corpus and status planning complement each other (see Figure 3.1). As Fishman puts it, they are “Like the two sides of a coin; they are obviously both necessary and co-present and any effort to influence one of the two cannot be engaged in without having necessary implications and repercussions for the other” (80). Decisions concerning status planning and corpus planning are often influenced and determined by both individuals (e.g., linguists, researchers, teachers, etc.) and formal organizations and institutions (e.g., universities, schools, professional associations, printing and publishing houses, etc.). They also require official intervention (by governments and their representatives) in order to be effective.
Cooper (109) adds a third distinction to status planning and corpus planning, acquisition planning (sometimes known as language-in-education planning), which involves only the educational sector of the society. It refers to ways to ensure that the chosen language is taught and learnt in school (Wright 1) and to make teaching and learning that language easier and more attractive. Acquisition planning is substantially different from status planning and corpus planning in that, while status and corpus planning penetrate many sectors of society as they are a function of the government, acquisition planning affects only one sector of society. It involves only the formal educational structure and does not affect even other educational and training activities in areas such as economics, tourism, and the military. It involves, for instance, curriculum design, teaching methodologies, evaluation, and testing. Kaplan and Baldauf (125) define five steps that form the major considerations for acquisition planning and implementation:

1. curriculum (the education sector has to decide on the language(s) to be taught within the curriculum);
2. personnel (it has to prepare and train, at both pre-service and in-service stages, teachers who will teach the new language elements);
3. materials (it has to prepare the language teaching materials included in the curriculum and to determine the methodologies that will be implemented);

4. community (it has to determine the level of students who will learn the language and get parental and community support for teaching the planned language, including financial sources for funding and maintaining the educational program in the short and long term); and

5. evaluation (it has to define the assessment procedures to measure student, teacher and system performance).

LP also exists at three levels, as distinguished by Bamgbose (109): official LP, educational LP, and general LP. Official planning (OP) is concerned with which language (or languages) to use at the government level; educational planning (EP) deals with the use of language in different kinds of academic institutions; while general language planning (GLP) refers to language use in general communication, business and contacts with foreigners (ibid.). EP plays a central role in the LP process because “it is the single most important vehicle of language policy” (Grin 43). Further, the acquired level of language skills is usually the outcome of the selected language education policy.

3.5 Terminology Planning

For many years, TP has been regarded as a part and parcel of LP. Many studies on LP (e.g., Bamgbose 1991, Antia 2000, Drame 2009) explicitly mention the salient role of terminology in LPP. The major concern of LP is the development of a lexicon (other aspects, such as standardization, might be implied). However, TP is not just about language; it consists not only of linguistic representations, but also non-linguistic representations such as graphic representations, formulae, numbers and signs. It is strongly associated with economic, political,
and development aspects. While the focus of LP is the conscious manipulation and development of a linguistic entity to improve communication in society or a language community at large, TP may be language-independent. Its objective aiming at the improvement of communication in a specific domain or application (Drame (87)).

In recent decades, while constituting a major element in LP processes, TP has become increasingly regarded as a separate entity that has recently emerged in many parts of the world at different levels (e.g., international, national, regional, local or institutional) and in many professional domains (e.g., information technology, chemistry, biology, physics) where terminological data are an indispensable element of such domain-specific knowledge and information. It has been regarded as a central element for communication and knowledge transfer that urgently requires discussions, particularly in multilingual regions such as ANAC.

A distinction should be made in this regard between domain communication and technical communication. While the former is used for specialized and professional subject fields such as scientific-technical subject fields in spoken and written language, the latter is used in the preparation of documents that contain special language primarily in written language. TP, therefore, develops terminology according to the needs and requirements of domain communication (Infoterm 11). Since terminology is the vehicle of scientific-technical communication, each domain needs to systemize, document and disseminate and then standardize its own terminology. Then, it would be possible to develop communication.

Language and terminology planning are two complementary concepts under a broader concept represented by the term ‘communication planning’ (CP). Both are regarded as separate but related activities with the goal of improving communication. They both reflect many criteria including, but not limited to, political context and the historical genesis of language use, the
socio-economic situation and demographic and cultural factors. CP takes a much more important role as it includes both LP and TP processes. Adopting a schema from Infoterm (16), Drame (89) proposes the following representation to visualize the interrelation between the three fields:

![Diagram showing the relationship between LP, TP, and CP.](image)

Figure 3.2: The relationship between LP and TP

Figure 3.2 gives an overview of the relation between these concepts. It illustrates terminology standardization as a superordinate term to language and terminology planning, whereas the reality is that terminology standardization is a central issue in the process of general CP. It involves selecting, among variant synonyms, the most adequate and correct one (Lauren et al. 287) according to standard criteria in order to create unified and standard terminology. Another fact which should be taken into consideration in the distinction between LP and TP is that the first is usually monolingual. It involves immediate issues such as language education (not mentioned in Drame’s representation). In contrast, TP is often multilingual and involves several specialised disciplines, for instance, when multilingual dictionaries or databases are
produced. Moreover, the degree of interdependence between LP and TP varies from one language community to another according to the specific circumstances of that community (ISO 29383).

3.6 Who Plans Terminology?

The question of who plans terminology has been raised by Antia in 2000 in his book *Terminology and Language Planning: An Alternative Framework of Practice and Discourse*. He suggests that terminology planning can be viewed as a bottom-up or top-down process. Antia emphasizes that most developing countries typically follow the traditional policy approach (top-down approach), which means agencies in charge of TP do not necessarily listen to the practitioners in the field, whereas in most developed nations a cultivation approach (bottom-up approach) can be observed (Antia 15). It seems that Antia adopts this view from Neustupny (1970) who suggests that LP and TP in developing countries is characterized by what he calls ‘policy approach’ ‘top-down approach,’ whereas in developed countries it is characterized by ‘cultivation approach’, i.e., a ‘bottom-up approach.’ The following section provides an example of the policy approach by introducing how the process of LP in Libya was formulated by the government in accordance with its political objectives.

3.6.1 Top-down vs. Bottom-up Model

As noted, in many developing countries, language policy is generally top-down (see Figure 3.3) and involves a procedure through which a linguistic group is manipulated by policy makers, usually politicians who do not leave anything to individuals to decide. Accordingly, terminology as a discipline is governed and influenced by the language policy of the government. It determines the number of languages to be developed as official languages (e.g., monolingual, bilingual or multilingual) and thus functional languages for terminology. Many
decisions regarding LP, TP and CP are initiated and formulated by government agencies and largely underpinned by political government objectives. For example, in Libya decisions related to LP have been made by the government and its representatives, who hold the highest positions on the administrative pyramid and who dominate the economy. They lack the competence to make policies related to language or language strategic planning. They even lack basic forms of knowledge in, for instance, domain knowledge and terminology knowledge\textsuperscript{11} that can improve the quality of language and terminology output. Policies tend to come directly from the authority, without any consideration for independent sources of expertise. Since Libya obtained independence in 1954, it has been a national priority to promote Arabic as a component of national identity and to deny the existence of multilingualism and linguistic variety and to promote a \textit{monoglot ideology}, a phenomenon defined by Silverstein (8) as an “Ideologically configured belief that a society is monolingual coupled with a denial of practices that point toward factual multilingualism and linguistic diversity.” No efforts have been made to consult professionals and specialists in areas such as education or sociology for their views on LP.

\textsuperscript{11} Infoterm (2005) identifies three basic forms of knowledge, which are applicable to all languages: linguistic knowledge (i.e. generic and language-specific knowledge); domain knowledge (i.e., conceptual knowledge of the contents of a particular field) and terminology knowledge (i.e. knowledge of terminology work methods and principles).
TP at national and regional levels is defined as a public strategy formulated at the level of political decision-making in a country or in a more or less autonomous language community with the aim of developing emerging terminology for any set of purposes (Infoterm 4). Figure 3.3 above illustrates the hierarchal relationships among language and terminology stakeholders. The arrow indicates the downward direction where the influence as well as information flows from one level down to the next.

However, the model with its emphasis on top-down requirements seems unsuitable for the development of languages in general and terminology work in particular. The fact that LP and TP ultimately require political decisions does not imply that societies cannot have any kind of influence on their linguistic environment. Grin calls this public policy, and Fishman states that these efforts are frequently overlooked as topics in the field of LP and TP:
The role of individual language planners has also been slighted in our deliberation. Many languages have benefited from the contribution of particularly charismatic and authoritative advocates, innovators and normifiers. We really know all too little about more than a mere handful of them, and as a result, we really lack any theoretical approach to their success and failure. (423)

For example, private operators may run television channels to promote a minority language or to improve foreign language skills among a school population without state intervention. Fishman argues that “government should not intervene, and that maximum welfare [feeling of safety and satisfaction in one’s identity] will automatically flow from the decentralized actions of people” (individuals, firms, third-sector organizations.)

Figure 3.4: Bottom-up model
Language Terminology planning, however, should ideally be a networking process in which different interest groups are involved, rather than a strictly top-down or bottom-up process. It must be bi-directional; from bottom-up, to bring the topic of terminology to the political or management agenda, as well as top-down, to ensure the support and involvement of those who are supposed to benefit and contribute to it. Drame emphasizes that “Both approaches influence each other and a strong governmental policy along with economic development may be a catalyst for stronger cultivation” (85).

Terminology planning takes place at different levels within a society or community. It depends on a large number of stakeholders from different communities and sectors, but it is typically designed by experts in the field, either individuals or groups of people (translators, researchers, language specialists, LSPs, etc.) who are aware of existing terminological shortcomings and are motivated to change the situation. Subject-field experts have a central role, as they work in both directions towards official preparation and formulation and implementation of changes. Lara (98) supports this view, as he claims that “terminology is a creation of the specialists in each field and not of linguists or terminologists. The latter organize, advise specialists, solve linguistic problems, research the present state of certain areas, create databanks, but they do not originate terminologies.” However, in many cases, especially in the developing world where concepts (which are the basic units of subject-field-related cognition, knowledge and communication) are needed to be developed in different linguistic environments, there is a higher degree of reliance on linguists and terminologists in order to find harmonious equivalents. For instance, Ereilt and Saari (1991, cited in Draskau 8) explain in their discussion of the terminology situation in Estonia that “the number of linguistically-trained terminologists has never been considerable in Estonia, yet they manage to retain control of LSP work.” Similarly,
Drame maintains that “[Terminology planning] is primarily a management activity. A central and often pro-active role within this activity is assumed by terminologists and other language experts; he or she is often the first person to recognize a need within a system” (100).

An adequate model of LP and TP should accommodate not only governmental decisions, but also several levels of non-governmental decision-making and implementation. Sager (210) maintains that “the ultimate choice of terms [...] must, however, rest with the user, as it is the user who originally invents or develops and names the concept.” The question is: how can we achieve realistic LP and TP, and through what process? How can we create an integrated top-down and bottom-up methodology that is liberated from political influence? What are the social, political and economic conditions that might make this possible? What are the consequences of this integration?

3.6.2 Integrative Model

LP and TP are in fact multifaceted phenomena; they are basically controlled and constructed from above, but cannot be understood unless also analyzed from below, taking into account the ordinary persons who are the object of the above actions from governments, as well as their spokesmen and activists for national movements (Hobsbawm 1). Decisions influenced by these activities are determined by both individuals and formal organizations and institutions. LP and TP also require official intervention (by governments and their representatives) in order to be effective. LP and TP should be an interdisciplinary, multi-directional process in which various external factors have influence and thus have to be considered. Neglecting these factors will endanger the success of this policy. Concerns include demographic factors, cultural factors, ethno-linguistic factors and socio-ideological factors. Some of these factors that may influence
policy decisions cannot, however, always be predicted (e.g., political and economic changes).

The following model is proposed by Antia (11):

![Figure 3.5: Integrative model](image)

This model (Figure 3.5) indicates some measure of coordination and suggests the limitation of both top-down and bottom-up models. It reflects the interdisciplinary nature of terminology work as a network process in which different interested groups (stakeholders) are involved, rather than a strictly top-down or bottom-up approach.

### 3.7 Term Formation

In order to understand term formation, it is first necessary to distinguish a term from a word, although it is hard to draw a clear borderline between the two. A term may consist of one simple word or a complex phrase—‘multi-word term’—with specific morpho-syntactic and
morpho-semantic features\textsuperscript{12} that may be unknown in general language, but that can be specific to certain domains (Infoterm 19). Term formation should not be a random or arbitrary process, but rather a conscious activity which differs from the arbitrariness of general word formation because term formation requires a greater awareness of pre-existing patterns and models, and of the social responsibility for facilitating communication and transmitting knowledge. (Sager 25)

Term formation follows different patterns and occurs in different environments, such as on the part of governmental institutions, professional organizations, universities, and among individual translators and writers. Ideally, however, LPP experts such as terminologists, linguists and translators play a crucial role, since they have the insight into and expertise in the problems of their domain. ISO 704\textsuperscript{13} (22-25) lists basic general principles governing the formation of terms, and these may be applicable to all languages: transparency (i.e., the term’s meaning is visible in its morphology), consistency (i.e., terms must be coherent with and integrated into their concept system), appropriateness (i.e., terms must be familiar to a language community and should not cause confusion), linguistic economy (i.e., terms should be as concise as possible), derivability (i.e., terms should allow appropriate derivations), linguistic correctness (terms must conform to the morphological, morpho-syntactic and phonological norms of a given language) and preference for native language (i.e., native language expressions should be given preference over direct loans, except in domains where the use of foreign forms is inevitable). Another principle

\textsuperscript{12} A \textit{Multi-word term} is a term that is composed of more than one word. Many examples can be found in the domain of computers and Internet. To name a few, Dynamic HyperText Markup Language (DHTML), Dynamic Host Configuration Protocol (DHCP) and File Transfer Protocol (FTP).

\textsuperscript{13} ‘ISO 704 Terminology Work, Principles and Methods’ is an ISO (International Organization for Standardization) standard, which defines the basic principles and methods for preparing and compiling terminologies both inside and outside the framework of standardization, and describes the links between objects, concepts, and their terminological representations (www. iso.org)
that governs term formation might be added to this list—*universal acceptance* (i.e., if terms are universally accepted in many languages and do not pose grammatical or understanding problems for a language, they might be considered with slight changes in pronunciation and/or spelling, depending on the conventions of that language).

In theory, the terminology development process can be described as comprising four different stages: term formation, term standardization, term dissemination, and term documentation.

![Figure 3.6: Stages of the terminology development process](image)

The problem with Arabic terminology formation in general is that term formation and production is scattered and unsystematic, which leads to a variety of synonymous Arabic terms for the same concepts. This variation is most acutely felt in the arena of specialized terminology, such as computer and Internet terminology. Besides the example of the term *computer* described in 1.2 (Research Problem), another example is the English term *laptop*, which is translated into Arabic as *كمبيوتر محمول* (kombutr mahmol), *المحمول* (almahmol), *بورتابل* (bortabl), and *لاب توب* (labtob). Rendering the term as *محمول* (mahmol) may be found in North African dialects, particularly Tunisia, Algeria and Morocco. The translation *كمبيوتر محمول* (kombuter mahmol) and the loan *لاب توب* (labtob) are seen in writing in Egypt and the Middle East, whereas *بورتابل* (bortabl) is only seen in Algeria and Morocco’s dialects where French is the dominant colonial language. It is worth noting that rendering the concept as *بورتابل* (bortabl) returns to the original French term *potable*. The direct borrowing of the English term as *لاب توب* (labtob) is most preferred and commonly
found throughout the entire Arab World. Since direct loans are often the only options that exist for certain specialized discourses, such as computers and the Internet, direct loans are occasionally adopted in the use of MSA by official academies (see Benabdi 1986 and Wilmsen and Youssef 2009). Many scholars attribute this variation to the dynamic nature of the domain of computers and the Internet. This will be discussed in the following section before reviewing in further detail the different patterns put forward by Arabic scholars to deal with the current status of computer and Internet domain loss.

3.7.1 Domain Loss in the Field of Computers and the Internet

It is unquestionable that computers and the Internet are part of everyday life; they are used in universities, schools, homes, markets, and banking, both online and onsite. Hence the use of computer terms is far-reaching and not limited to merely one social or professional group. Terms from this domain are widespread among specialists and non-specialists alike. Therefore, consistent treatment and use of computer and Internet terminology throughout countries, organizations, individuals, and various social strata to improve communication and integration between production and documentation is becoming indispensable for Arabic to function effectively (either internationally or locally) as a scientific and economic language. As Drame (2009) puts it, societies cannot achieve any significant level of modernization and development without systematic terminology use and treatment. Therefore, assessing the terminological variations in formal written MSA (the main focus of this study) is becoming a must in order to counter the current domain loss. Domain loss is defined by Laurén et al. as “Loss of ability to communicate in the national language at all levels of an area of knowledge because of deficient further development of the necessary means of professional communication” (5), particularly in the dynamic domain of computers and the Internet. One of the distinctive features of this domain is its dynamic nature. For example, until recently, terms like ipod, ipod touch, ipod apps, ipad,
*ipad apps*, and so on were practically unknown, but they are now widely used to refer to their well-known corresponding techniques and systems.

Domain loss is defined by Laurén et al. as “Loss of ability to communicate in the national language at all levels of an area of knowledge because of deficient further development of the necessary means of professional communication” (5). The nature of Arabic language (see Chapter Two) has resulted in the current trend of Arabic toward ‘domain loss,’ i.e., loss of Arabic capacity to generate new terms and expressions due to the hegemony of foreign languages and the insufficient mastering of MSA as a lingua franca for the majority of the population. The fact that MSA is taught at elementary and secondary schools almost as a second language (see Chapter 2), which in turn resulted in the phenomenon that large sectors of the Arabic public do not command MSA sufficiently, has made it difficult for Arabs to develop successful scientific and technical terminologies, and it has forced even experts to use foreign terms for communication in scientific and technical domains. It is worth noting that this difficulty is not only due to the complex nature of MSA. The fact that Arabic and English are incongruent languages, linguistically and culturally, is another main factor. In his textbook *Towards a Scientific Translation* (1964) Nida points out this challenge:

If, however, the translation of scientific texts from one language to another participating in modern cultural development is not too difficult, it is not surprising that the converse is true that translating scientific material from a modern Indo-European language into a language largely outside the reach of Western science is extremely difficult. This is one of the really pressing problems confronting linguists in Asia today. (223)
In such cases, when language undergoes significant domain loss with few terminological resources, terms are often created by non-specialists. As Laurén et al (7) put it, “the terminology of revitalized and less-widely-used languages tends inevitably to be the creation of middle-class language experts rather than that of specialists…terms are coined by linguists rather than specialists.” In the Arabic context and in the absence of systematic terminology planning, it is up to journalists or technical writers who face the problem of not finding an equivalent in the TL for a particular SL term to choose the most suitable way to render it *ad hoc* by coming up with their own loan translations from their colonial languages without following any systematic criteria for forming terms in Arabic.

### 3.7.2 Methods of Term Formation

One of the most important and urgent tasks for any developing language in the spheres of science and technology is the deliberate effort to cope with foreign terms that are created every day, by forming and introducing the most appropriate counterparts of such terms. With the colonial period (1830-1950), which made foreign languages (mainly French and English) the languages of government administration in many Arabic countries for an extended period of time and established them as the medium of communication and learned science and technology, new challenges emerged. Arabic, particularly MSA, has to compete with these well-established languages. Indeed, there is competition taking place among these three languages, as French is competing with English, Arabic is competing with both of them, as well as with the indigenous languages, and the indigenous languages themselves are threatened by Arabisation policies. This competition is the outcome of that long history of language contacts.

As a result of these contacts, new scientific and technical terms and concepts started to penetrate Arabic. This phenomenon has drawn the attention of Arab scholars to the need to
create and introduce new methods of terminology creation in Arabic in order to handle the flood of foreign scientific and technical concepts. The following discussion will look into the main methods of term formation that have been developed by Arab scholars in order to suit the specificities of Arabic as a linguistic system. These methods are: creating new terms (such as derivation, compounding and abbreviation) and trans-lingual borrowing and adaptation (such as direct loan and loan translation). Either the term is just taken in its original form (with some modification for pronunciation and spelling) or the semantic components of the term are translated to form new compound terms.

3.7.2.1 Creating New Terms

3.7.2.1.1 Derivation

The first method of term formation favored by Arab academia is derivation. Derivation ‘الاشتقاق’ al-ishtiqiq is a process that involves forming a new term by adding one or more morphological elements (affixes) to a root or a word (ISO 704, 28). It is considered one of the main techniques employed by many ancient Arab scholars (e.g., Al-Marghani and Ibn Faris) to improve and enrich Arabic vocabulary with new terms. As a Semitic language, Arabic is equipped with a powerful and flexible word formation system14 (Ibrahim 5). In Arabic, as in English, affixes (which depend on the attachment of the three morpheme types: prefixes, infixes, and suffixes) can produce a new word, as in تعلّم تعليمي education, educational. Most words in English are simple roots, while in Arabic derivation plays a much more pervasive role. It differs from inflection in that unlike inflectional morphemes, derivational morphemes form new words either by changing the meaning of the base to which they are attached (e.g., like, dislike) or by

---

14 Different patterns of word composition may be found in Semitic languages such as Hebrew and Arabic, which are characterized by discontinuous morphology (Shimron 225). They involve structures in which the root is modified and do not involve attached morphemes.
changing the word class that a base belongs to (e.g., happy, happiness). In Arabic almost all words are derived from roots by the addition of affixes, thus it is known as a language of derivation. Three main forms of derivations are distinguished in Arabic (see Elmgrab 2011):

(i) Simple derivation ishtiqaq sagheer, in which the radical consonants are not altered but are derived from and built upon. The examples in Table 3.1 show the possible forms that can be derived from the three core consonants (ع, ل, م), which carry the core meaning ‘to know’. Typically, the root employed in Arabic to represent such derivations is the trilateral فعل fa‘al (which means did). There are many derived forms of this trilateral, as indicated in the following table together with their correspondent derivations from the radical ع ل م:

<table>
<thead>
<tr>
<th>Trilateral fa‘al</th>
<th>Transliteration</th>
<th>Radical a’lima</th>
<th>Trilateral fa‘al</th>
<th>Transliteration</th>
<th>Radical a’lima</th>
</tr>
</thead>
<tbody>
<tr>
<td>فعل fa‘al</td>
<td>علم af‘al</td>
<td>أفعال اعلام</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>فعل afa‘lal</td>
<td>علم مقاطع</td>
<td>مفعال ماعلیم</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>فعل fa‘al</td>
<td>علم فعلاء</td>
<td>مفعل علماء</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>استفعل istaf‘ala</td>
<td>علم فعل</td>
<td>معدل معلم</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>تفاعل tafa‘al</td>
<td>تعاليم فعال</td>
<td>fa‘‘al علمیم</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>مفعل maf‘ül</td>
<td>معلوم فعل</td>
<td>fa‘eel علمیم</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>فعل fa‘ül</td>
<td>علم مقاطعة</td>
<td>moufa‘ala معالمة</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Derivation methods

Table 3.1 above shows that 14 words were derived from the three-letters root ع ل م (radical a’lima). Each form can be transformed into one of the 14 possible verb forms ‘وزان’ awzaan (trilateral fa‘al). For example, علم a‘lam was formed by prefixing ا alf and dropping the vowel
of the first letter. Likewise, تعلّم ta’lam was formed by prefixing ت ta and infixing the أ alf after the first letter. Each derivational form has a basic meaning associated with the general meaning of the root (ا’lima).

(ii) Wider derivation ishtiqak kabeer, which involves a change in the position of the root consonants and retention of the original meaning. It is also called إبدال root derivation or ibdal\(^\text{15}\) substitution. An example of this derivation shows the following verbs:

- نبذ nabaḍ to reject
- قطع qata’a to cut
- قطم qatama to cut
- قطف qatafa to cut

These forms are structurally parallel and differ from each other merely in their final letter. This difference makes the sound of each word very similar to that of the others, thus the semantic proximity is preserved. Some scholars (e.g., Al-Lugawi 1962) have attributed this approach to Arabic dialectical differences.

(iii) Another form of derivation that has been newly introduced as one of the methods of producing new terms is ishtiqaq bi-atarjamah (literally, derivation by translation). This is also called ishtiqaq ma’nawi (literally, derivation of meaning). El-Mgrab (495) argues that although circumlocution is widely used in neologisms, it has led to the problem of a multiplicity of terms due to the fact that there are no exact rules used in translating foreign terms. Translators often invent their own arbitrary translation that does not follow international or even Arabic models and rules. For example, the word Internet is translated into Arabic as الشبكة الدولية

\(^{15}\) Ibdal in Arabic means substitution or replacement in the root of words, for example the two Arabic words لقاطa and لقافa have the same meaning (to pick up), but the ‘t’ in the first was replaced by ‘f’ in the second.
Ashabaka Adawleya *international network*, AShabaka Al-ankaboteya *spider network* and Ashabaka Al-baineya *internetwork*. Since there is no Arabic standard source of translation that effectively introduces proper terms, variant forms occur in different strata of Arabic society.

It is worth mentioning in this regard that derivation in Arabic can be carried out by changing vowels. The Arabic vowel system differs from English or any other western language.\(^{16}\) This feature makes Arabic orthography homogenous, which means that one word may have several pronunciations and several meanings and functions; any changing in short vowels usually leads to a change in the lexical meaning (Abu-Rabia 231). For example, the trilateral radicals (ع ل م) can mean ‘science or knowledge’ if pronounced as (عُلْمُ) e’lim, a *flag* if pronounced as (عَلَمُ) ‘alam, *known* if pronounced as (عُلِّمْ) ‘ullim, *to be taught* if pronounced as (عُلِّمَ) ullim, etc. The correct pronunciation and meaning of this word can be obtained by means of vowels. Since including the short vowel symbols is not the practice in most Arabic texts, skilled Arabic readers usually place the word in its context in order to intuit the correct lexical meaning.

It should be noted in this regard that all the above derivations are formed from verbal roots. However, many forms can be also derived from abstract nouns by adding a final suffix to nouns in order to coin new terms. Noun derivation is now widely used as one of the practical ways of coining new terms in Arabic, e.g., اللوغرثمية al-lughatmiah *algorithmic* is derived from اللوغرثم al-lugharitm *algorithm*.

\(^{16}\) Arabic has 28 letters, which are all consonants. Some of them (أ، و، ي، a, waw, ya), however, can be considered as long vowels in some cases, and they are part of the Arabic alphabet. Unlike English, short vowels in Arabic are not part of the alphabet. They are specific signs تشكيل *tashkeel* added to take the form of tiny diacritics marks placed above or beneath the (consonant) letters. In Arabic, there are three main short vowel signs (ا، ى، ع، a, i, u), as well as other signs حركات *harakaat* that indicate the silent sounds and the prolongation and the double pronunciation of a consonant.
3.7.2.1.2 Compounding

A second method of term formation favored by Arab academia is *compounding* (also known as *blending*), which involves combining existing words or word elements to create a new form that contains two or more roots but designates a single concept (ISO 704: 28). This compounding process is very common in languages like English and German (for example, *bookcase*, *shoemaker* and *wastebasket*), but much less common in languages like Arabic. In general, the most common form, according to Sager (13), is actually the linkage of two independent morphemes, which may or may not be independent words on their own. Traditionally, this method is known among Arabic scholars as *النحت anahat* sculpturing. It has been used recently for coining new terms, especially technical and scientific terms—e.g., the term *كهرباء ساكنة kahraba sakeenah* electrostatic is formed by conjoining the word *كهرباء kahraba* electric and the word *ساكنة sakeenah static*. It is sculptured into Arabic as *كهراكدة kahrakada* where the first two syllables in the *كهرباء kahraba* and part of the second syllable in the word *ساكنة sakeenah* were combined. Another example is the term *تحتربة tahturbah subsoil*, in which the radical *سول soil* in translated as *تربة turba* and the prefix *سوب sub* is translated to the preposition *تحت تحت tahata under*. Then the preposition *تحت تحت tahata* is reduced to its first syllable *تح táh* and attached to the radical as a prefix, which gives *تحتربة subsoil*. Al-Mawred Dictionary goes further in this regard and combines three or more independent morphemes, which in many cases results in distorted structures that violate Arabic morphological rules. Consider the following examples:

| شللاليتلي | shalaliufaily | poliomyelitc |
| تحولوينتري | tahawaltablury | paramorphic, paramorphous |
| اختلافينظري | ikhteelahifanzyary | paraltic |
| سفرمحيطي | safarimohetey | oceangoing |
This type of term formation, according to Ryding (50), has become more prevalent over the past two or three decades with the need for rapid translation of technical and computational terms from European languages into Arabic.

Arabic philology distinguishes two main forms of compounding:

(i) Full compounding, by which two existing words are joined to form a new word. For example, لوحة مفاتيح lawhat almafateeh keyboard, سطح المكتب saṭah almaktab desktop, لوحة التحكم lawhat altahakum control panel, القرص الصلب alqurs alṣalb hard drive. Sometimes more than two words are joined to form a new term, as in قرص ذو كثافة عالية قرص ذو كثافة عالية high density disk and اختصار لوحدة المفاتيح ekhteešar lawhat almafateeh keyboard shortcut.

(ii) Partial compounding, by which only parts of two or more words are joined to form a new word. This is typically accomplished by taking only the beginning of one word and joining it to the end of the other word. For example, أوروبا وأسيا Europe and Asia are compounded into أوراسيا euorisia, أفريقية وأسيا African and Asian are compounded into أفراسيوي afroasyawi.

3.7.2.1.3 Abbreviation

A distinction should be made between shortening (clipping) and abbreviation. Clipping is a word formation process which consists of truncating a part of a simple term (ISO 12620: 6). It usually denotes the subtraction of one or more syllables from a word. It may occur at the beginning of a word, as in: ‘telephone’ phone, ‘photograph’ photo, mathematics ‘math’, gymnasium ‘gym’, flu ‘influenza’, or — which is most frequently the case — at the end of a word as in: ‘hamburger’ burger, French fries ‘fries’. Abbreviations, on the other hand, are terms
resulting from the omission of any part of the full form while designating the same concept (ibid: 7). The full forms are generally represented by the initial letters of their main lexical constituents, as in 'أف بي أي' FBI (Federal Bureau for Investigation), 'بو أس أي' USA (United States of America), 'دليو اتش اور' WHO (World Health Organization) and so forth, but in many cases the letters represent elements in a compound or parts of a word, e.g., 'تي في' TV (television) and 'جي أي' GHQ general headquarter. Abbreviations are commonly used in written language, and many written abbreviations have no spoken equivalent, such as 'كم' kilometer, 'بص' P.O. Box, 'سنتيمتر' cm centimeter. However, the use of the full form of expressions such as ‘P.O. Box’ is more acceptable than abbreviation 'صندوق بريد' Post Office Box. Although many of these abbreviations such as km and cm are widely international in form and are identified as ‘Standard International Units’ (IS) because they have the same or nearly identical orthographic or phonemic forms in many languages (ISO 12620: 6), they usually reflect Latin, Greek, French or English origin.

The phenomenon of abbreviations in Arabic dates back to 175 A.H. (797 A.D.), when Ibn-Faris discussed in his book *Measures of Languages* how to deal with abbreviations and acronyms 17 in Arabic. His theory was that terms consisting of more than one word are formed by taking the first letter in every word. Every letter equals a distinctive Arabic word (see Al-Khalil 1980). In Arabic, abbreviations are commonly used in religious texts such as the Qur’an. Twenty-nine surahs (chapters) in the Qur’an begin with abbreviations; to name some: ha mim ‘ح’, kaf ha a’an saad ‘كﮭﻊ’، ‘اﻟﻢ’ In general, however, abbreviations are not as widely used in Arabic as in other languages such as English. However, they have been increasingly used during the recent decades in response to the linguistic necessity imposed by technical development as well as the hegemony of English-language style. Many abbreviations (e.g., HTML 'لغة إعداد النص')

17 Acronyms are distinct from abbreviations in that an acronym is pronounced as a word. Thus, there should be a pronunciation harmony between its parts as in NATO (North Atlantic Treaty Organization), NASA (National Aeronautics and Space Administration) and SCUBA (Self Contained Underwater Breathing Apparatus).
SGML (لغة الترميز العمومية الأساسية) and HTTP (لغة نقل النص) are transcribed in Arabic in their Latin form (using English characters), and their translated full forms are added in parentheses for clarification. In many other cases an English abbreviation is translated into Arabic through a whole statement or phrase. For example, the abbreviation ‘W3C’ (World Wide Web Consortium) is rendered into Arabic as جمعية تحديد معايير لغة إنشاء الصفحات (Determining Web Standards Consortium). Similarly, the following examples are also approved by academies:

- **CET** توقيت وسط أوروبا (Central European Time)
- **IMF** صندوق النقد الدولي (International Monetary Fund)
- **PC** حاسوب شخصي (personal computer)

“Because of their specific mode of formation, abbreviated terms require a special treatment in Arabic as individual letters do not form meaningful units in that language” (Yousif 314). In Arabic, however, abbreviated terms are often created on an ad hoc basis by translators and technical writers. Many translators and writers create terminology on a daily basis, and they do not even have basic skills with regard to word-formation principles. For instance, the term ISBN (International Standard Bibliography Number) has been abbreviated into a non-meaningful Arabic unit as تدمك ‘تدمك’. Moreover, variation frequently occurs as a result of adaptation from different sources, even for abbreviations where there are established concepts in MSA. For instance, the two terms ‘أيدز’ AIDS from English and ‘سيدة’ SIDA from French are both approved by Arab academia. Another example is the abbreviation ‘ناتو’ NATO adapted from English and ‘أوتان’ OTAN from French. Abbreviations may also be combined with numerals to generate alphanumerical constructs such as ‘i18n,’ which stands for ‘internationalization’; 18 represents the 18 letters between the initial ‘i’ and the final ‘n’.
3.7.2.2 Translingual Borrowing and Loan Translation

3.7.2.2.1 Loan Words (direct loan)

Another widespread way of creating new terms is direct loan words. A loan word (also known in Arabic as kalima dakhela كلمة دخيلة intrusive word) is a word directly taken into one language from another with little or no translation. Throughout its history, Arabic has adopted a vast number of loan words from other languages, including video فيديو bus, faks فيدئو، باص، فيديو، فاكس from English; etiquette إتيكيت، رجيم، برستيج prestige from French; and esperto espíritu، أوبرا، ubera opera from Italian.

In spite of the many efforts to ensure that terminology formation is undertaken using standard resources such as derivation and compounding, many technical terms have been introduced into Arabic as direct loans. As a consequence, the Coordination Bureau of Arabisation (CBA), which was established in 1967 in Morocco under the auspices of the Arab League Educational Cultural and Scientific Organization (ALESCO), recently (in a conference held in 1982) acknowledged the inevitability of borrowing for adapting foreign technical and scientific terms into Arabic.
In many cases these loan words are favored by Arabic speakers over their native counterparts (e.g., تلفون telephone is preferred over هاتف hateef, موبايل mobāl mobile is preferred over هاتف جوال mahmūl or jawāal, راديو radio is preferred over إذاعة مسموعة āḍa’a masmoo’a). There are many factors that might promote this preference for loan words, such as foreign language prestige (discussed later), but the adoption of loan words is primarily due to the fact that MSA had not established standard Arabic terms for these concepts. Even when there is an established standard term, as in the case of حاسوب computer, it is used only in official documents. Therefore, the common choice will be the simpler alternative ‘loan word’. These words are most noticeable in the arena of specialized terminology, especially for technology such as information technology, aeronautics, and the automotive industry, where standard Arabic terms are lacking. These loan words can hardly be comprehended by average educated Arabic speakers from different dialect areas. For example, colloquial automotive terms, which were originally adopted from Italian, are the only existing option for Libyans to use. Table 3.2 shows some examples:

<table>
<thead>
<tr>
<th>English</th>
<th>Libyan Arabic</th>
<th>Transliteration</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clutch</td>
<td>فرسبيوني</td>
<td>frassūny</td>
<td>frizione</td>
</tr>
<tr>
<td>Transmission</td>
<td>كمبيو</td>
<td>kambyū</td>
<td>cambio</td>
</tr>
<tr>
<td>shock absorber</td>
<td>مازاتوري</td>
<td>marzātūry</td>
<td>ammortizzatore</td>
</tr>
<tr>
<td>ignition coil</td>
<td>بوبينا</td>
<td>būbena</td>
<td>bobina</td>
</tr>
<tr>
<td>big end bearing</td>
<td>برونزيني بیلا</td>
<td>brānzanby la</td>
<td>bronzina di biella</td>
</tr>
<tr>
<td>Carburetor</td>
<td>کوربرتوري</td>
<td>karbaratūry</td>
<td>carburetore</td>
</tr>
</tbody>
</table>

Table 3.2: Terms of Italian origin in Libyan Arabic
The English term *clutch*, for which there are many equivalents in MSA (e.g., ﻗﺎﺑﺾ رعات, ﻗﺎﺑﺾ اﻟﺴـمـﻌـﻖ اﻟﺴـرـعـﺎت), is known as *klatish* in Jordan and Iraq, *دـبـﺮـﻳـاـج* diberyâh in Tunisia, Egypt and Morocco, and *فـﺮـﺳـﻰـنـﻲ* frassûny in Libya—loans from English, French and Italian respectively.

Loan-word variation frequently occurs even for terms for which there are established concepts in MSA, as a result of borrowing from different sources. In general, there are two main language sources for terminology (French sources in Morocco, Algeria, Tunisia and Mauritania, and English sources in Libya, Egypt and almost all of the rest of the Arab World). The pace and scope of term formation has depended largely on which of these languages is more dominant in a given area. Therefore, many Arabic technical and scientific terms are influenced by one of the source languages, either French or English. Consequently, terminological variation may occur as a result of loan words from these two different language sources, posing a potential communication problem for speakers from different parts of the region, as they may wrongly assume different meanings for each word. Another example (cited by Wilmsen and Yousif 205) is the term *privatization*, which is ﻫــﺨـﺼـىـ خـصـصـة khškhaša in Egypt but خوـصـصـة khawšaša in Morocco, or *recycling*, which is إﻋـﺎـدـادة إـعادة التدوير ائدة التدوير al-raskala, in Morocco.

Many loan words have been analyzed by Arabic linguists and grammarians in accordance with Arabic derivation patterns (mentioned above) and to determine from which sources new terms are produced. For example, the loan word تلفزيونون تلفزيون telfizyûn *television* has been reanalyzed into its root تلفز tlfz, and from this the noun مـتـلفـاز تلفاز telfaaz *television*, the participle مـتـلفـاز تلفاز mutalfaz *televised*, and the verb مـتـلفـاز تلفاز talfaz *televise* are derived.

To sum up, it can be argued that the status of loan words is uncertain in the host language, since they are derived from foreign words. Previous studies (Weinreich 1979) show that most loan words undergo some semantic, phonological, morphological or other alterations before or
even after they are accepted in the borrowing language. This is due to many reasons, such as the structure of the host language, inability to fully understand the meaning and usage of the original word and cultural and communication needs. These changes keep occurring as words are used repeatedly until they become an accepted form in the lexicon of the borrowing language.

2.7.2.2 Loan Translation

Loan translation (also called calquing), whereby a term from one language is translated element by element into the receiving language (Wright and Ludin 39), is one of the successful methods of term formation used by Arab experts to develop technical vocabulary in Arabic. It is used to designate new things and concepts. Here are some examples:

- The Arabic ﻥﺎﻃﺤﺔ ﺳﺤﺎب Natehat sahab, the German Wolkenkratzer (cloud scraper), and the French gratteciel (scrape sky) are all loan translations for *skyscraper*.
- The Arabic ﺛﻮﺣﺔ ﻧﻈﺮ ‘whhat nazr and the French ‘point de vue’ are loan translations for the English ‘point of view’.
- The English *wisdom tooth* is a loan translation for Latin ‘dens sapientiae.’

Over time, such loan translations become part of the Arabic phraseology and are no longer regarded as foreign. However, there is a feeling among Arab scholars that this tremendous volume of loan translations will corrupt the purity and destroy the spirit of the Arabic language. Bader supports this point, as he states that “while some loan translations usefully enrich the semantic repertoire and the expressive capacity of the borrowing language, others betray arbitrariness, inconsistency, and obscurity” (100). For example, both the English term ‘bus’ and its French counterpart ‘autobus’ exist as loan translations in Arabic, posing a potential communication problem for speakers from different parts of the region, as they may wrongly assume different meanings for each term.
Colloquial Arabic has always been open to loans from other languages and dialects. The adoption of terms from the local vernaculars is also usually acknowledged as a legitimate process for the construction of technical terms. Terminologists in ANAC countries consciously adopt colloquial roots for the coining of new concepts if no other alternatives are found (Benabdi 74). Currently, these colloquialisms are most prevalent in the arena of specialized terminology, especially in the arena of scientific discourse, where standard Arabic terms are lacking. Since people work in a vacuum and usually do not know whether someone else has already coined an equivalent for a term, it is usually up to the terminologist or translator who faces the problem of not finding that equivalent in Arabic to draw her or his own conclusions on the most suitable way to render the term in the TT.

2.7.2.3 Arabisation

Before I go any further, a distinction should be made between two terms that are constantly used interchangeably: ‘Arabization’ and ‘Arabisatin’. The first refers to the imposition and usage of Arabic in place of some other languages, either foreign languages such as French and English or indigenous languages such as Tamazight. Arabisation, on the other hand, which is considered as the second main method for creating neologisms when developing Arabic terminology, is defined as “the introduction of a foreign term into Arabic by an Arab who ‘boasts’ of his innate linguistic abilities, thereby conferring on this foreign term the grammatical status of a genuinely Arabic word” (Al-Iskandari 1935, cited in Elzeer 80). It is, thus, a conscious process of adopting a new term in Arabic in a way that preserves its foreign origin and at the same time safeguards the forms and spirit of the Arabic language. Although the two terms express different concepts, one Arabic term التعمیر ta’rib is used interchangeably for both of them. Ironically, the term ‘Arabisatin’ itself is ill-defined amongst those who are working on
the Arabisation process. It has been used by Arabic terminologists in three different senses. In one sense, ‘Arabisation’ refers to the utilization of Arabic as a tool of expression in all fields (Bahumaid 133). In another sense, ‘Arabisation’ only means translating a term into Arabic—e.g., French ‘terminologie’ into معطيات muṣṭalahaat. The narrowest sense of the term involves, as Khulusi (1982) put it, merely transliterating a foreign term according to Arabic characters—e.g., French ‘radar’ is arabised as رادار ‘radar’.

Since the early stages of the Umayyad era (the seventh century) and with the advent of Islam and the expansion of the Muslim empire, many foreign terms, especially Persian and Byzantine terms, continued to be widely used. This resulted in early works dedicated to Arabisation, such as al-Jawaliqi’s famous dictionary of Arabic terms Kitab Al-ma’arrab (1073). The early stages of Arabisation, however, mainly entailed transliteration using so-called ‘sound-letters’ (pronunciation in the original language). This is clearly evident in Al-jawhari, who simply defines Arabisation as:

"تعریب الاسم الاعجمی: ان تنقهو به العرب على منهجها" (Al-Jawhari 197)

Arabisation of foreign words should be in accordance with their pronunciation as well as with the traditions of Arabs. (My translation)

Elzeer (83) gives some examples: the Persian word برکار brkar compass was arabised as both فرکار and برکار brkar and the word جارو حارw arabis as صاروج šaruwkh rocket. The Greek letters ‘t’ and ‘k’ were consistently changed into ‘t’ and ‘q’ respectively. Contemporary scholars, however, oppose the introduction of sound-letters into Arabic. Al-karuri, for instance warns that there is no reason that Arabic should have to surrender its ‘identity’ in this way, especially since European languages do not do so and tend more towards transliteration (cited in Elzeer 84). Al-Karuri instead proposes the method of adaptation of foreign sounds.
In the introduction of his book *Terminology Science: Basic Theories and Scientific Applications* Al-Qasim (1) mentions that:

إنّ كلمتي "مصطلح" و "اصطلاح" مترادفتان في اللغة العربية. وهما مشتقان من "اصطلاح" (وجذره صلح) بمعنى "اتفاق"، لأنّ المصطلح أو الاصطلاح يدلّ على اتفاق أصحاب تخصّص ما على استخدامه للتعبير عن مفهوم علميّ محدد.

The words “muṣṭalah” and “iṣṭeelah” are synonymous in Arabic. They are derived from ‘iṣṭalah’ (its root ‘ṣalah’ means ‘agreed’) because ‘muṣṭalah’ or ‘iṣṭeelah’ indicate an agreement to use a certain term for a specific scientific concept. (My translation)

Many contemporary scholars argue against Arabisation as a method for developing Arabic terminology and instead favor other methods such as derivation and substitution (*ibdal*, see 3.7.2.1.1 above). Al-Ayli (357), for instance, calls for the restriction of Arabisation to proper nouns and names. Others (such as Madkur 39) even argue that the practice of Arabisation should be limited to what has been passed on to us by the eloquent speakers of ancient times—a number that has been estimated not to exceed one thousand words. The Academy of the Arabic Language in Cairo supports this restriction, as many of its decisions regarding Arabisation were related to Arabising proper nouns. While approving a large number of technical and scientific terms, its work has been largely focused on proper nouns in different categories such as personal names, names of animal and plant species and terms referring to breeds and genera.

To conclude, Arabization and Arabisation are two different terms, which refer to two separate but related concepts that seek for improvement of Arabic language and terminology. The failure of Arabic scholars to distinguish clearly between the two concepts is a not insignificant problem in developing Arabic terminology work. Redouane (199), for instance,
points out that the Arabisation process may heighten the problems posed by the coexistence of these varieties and the need to determine which should be used for what purpose.

3.8 Terminology Practice and Official Terminology Policy in the AW

In general, terminology as a science is a new field of investigation; no more than a handful of the 192 member states of the United Nations possess terminology policies (Antia 10). In the ANAC region in particular and the AW in general, academics and professionals have shown limited interest in terminology as a field of activity distinct from translation. There has even been considerable reluctance in relation to terminology work on the part of translators, who often view it as a time-consuming and boring process. Among other reasons, this can be attributed to the fact that constructing a terminology database, especially in Arabic, is relatively time-consuming and boring in comparison to constructing a translation memory (TM), especially with the ongoing ignorance of terminology management tools by Arabic organizations and individuals alike. Part of the problem is the Arabic language itself. Not all software applications support Arabic text, and great care must be taken when using Arabic in complex layouts. In fact, due to encoding/font difficulties and the bi-directional nature of Arabic, even simple operations, such as copying, pasting or sending an email, can become a nightmare when the browser does not support Arabic characters.

In ANAC, TP is often the outcome of individual efforts (linguists, translators, technical writers, researchers, etc.) who often work outside the framework of formal organizations, in spite of the fact that “language academics have done a great deal of work in the field of terminological expansion at both the theoretical and the practical levels, setting for word coining and Arabisation rules that they thought would help to control the impact of the absorption of foreign concepts into the Arabic language” (Elzeer 12). Their efforts consist only of listing terms. In
order to avoid misunderstandings, scientists almost always find themselves obliged to include foreign loan words in transcription next to their (own) word-for-word translations and sometimes even to write in the English words next to that as well (Ibrahim 5).

Much of the variation in terminology in the ANAC can be attributed to the ongoing competition and conflict among Arabic, French, English, and Amazigh. The widespread official realization that Arabic is the only politically, economically, and socially acceptable solution, the increased prominence of English and French as means of expression for modern science and technology and economic development, as well as the continued calls to use indigenous languages and their widespread use in radio broadcasts, magazines and newspapers have all resulted in competition and conflict among these languages for prestige and power. This situation has resulted in the coexistence of competing terms, so that their associated concepts are expressed by multiple individual terms. The translator has to determine whether Arab academies have provided official equivalents for those terms, as these academies are supposed to have the insight and expertise to produce translations and equivalent lexical items for most of the new and up-to-date terms and expressions in different fields of knowledge. The problem is, however, that language academies working in various parts of the ANAC typically respond slowly to the flood of new terminology, especially in the area of information technology. For instance, recent technological advances have produced new Internet expressions such as cloud computing, burnproof technology, and msipathedit, which are as yet too novel to be listed in any Arabic glossaries or dictionaries due to this slow response of language academies. Bahumaid (134) estimates that Arabic institutions in charge of coining Arabic equivalents for foreign terminology have dealt with only around 50,000 Arabic terms in the last sixty years. The inability to deal with the flood of terms has forced Arabic translators and technical writers to coin their own terms,
with the same term being rendered into multiple different Arabic forms. Consequently, much of the terminology work is performed ad hoc by newspapers and individuals who, unlike the academies, “do not have the luxury of methodically compiling data and meeting annually to make their pronouncements, but instead must meet regular deadlines.” (Wilmsen and Yousif 192) According to Hosny and Wahab (83), Arabic journalists have been the most active social group in introducing coinages into MSA.

To make matters worse, ANAC decision-makers neither encourage nor financially support terminology projects with the necessary funds to promote terminology work and to unify different translation equivalents for the same ideas and concepts. They are not aware of the usefulness and effectiveness of such a service. “[T]he huge price for starting up technical terminology in a newly computerized language for African or Indian languages” (DePalma and Joscelyne 30) seems to hamper the development of terminology activities in the region. One of the greatest difficulties is to persuade and convince the decision-makers, particularly politicians, to allocate more money to terminology development and management. In Morocco, for instance, “Financial support provided to translators by public authorities is weak, raising private funds is very difficult, and banks are reluctant to finance translation activities as well as other cultural activities.” (Achy 13). Part of the problem is the political situation in the region. The fact that there are six countries with many political issues among them poses a major obstacle to the process of standardization of Arabic terminology. Emery (85) supports this point when he states that:

Much of the duplication in terminology creation in Arabic can be attributed to geographical and political factors – the sheer size of the Arab world (and the problems of communications therein) and the concern of Arab states to build up
their own individual educational systems and develop their own political identities.

Likewise, Clyne (273) states that one outcome of having different political, administrative and ideological systems is the increase of language variability and difference, especially in the political domain. In Libya, for instance, the term أمانة amanah secretariat is used in place of the common term وزارة wazarah ministry, which is used in all other ANAC political systems. This gap in communication and coordination among institutions in the ANAC region has not only forced translators and technical writers to coin their own terms (as discussed above), but also has forced companies who want to conduct business in the Arabic world to translate their handbooks and product descriptions in order to present their products in MSA and to avoid unnecessary variants, corrections and miscommunication. As Ibrahim (4) observes,

The lack of any organized structures, government financial support or any private commercial sector that dedicates itself to this issue is the major reason why this task is being neglected in these countries. The lack of any uniform database for technical terminology causes considerable difficulties for multinational companies with regard to marketing their products on the Arabic markets, as well as for preparing appropriate translations for their product descriptions.

(Translated by KSU GER 61240)

Finally, we may conclude that one of the most debatable issues among modern Arabic linguists is the standardization and unification of Arabic scientific and technical terminology. It has recently attracted more attention than any other field of research (see Balry 1999). There is a widespread awareness among Arab scholars that the standardization of terminology contributes
to the quality of translation, editing, interpreting, services, dictionary compilation and subject-related communication, as well as reduces misinterpretation and misunderstanding. Nevertheless, no comprehensive attempt has been made towards the standardization of scientific and technical terms. An observer of the terminology discourse in ANAC will notice that those who are involved in terminology work do not agree on methodology, procedures or principles of terminology work (horizontal standardization), nor on terminology for specific subject fields (vertical standardization). Horizontal standardization is established by experts in terminology science, including linguists, language planners, professional translators, and so forth, whereas vertical standardization is the product of subject specialists of a given field. There is no deliberate TP in the ANAC. Much of the work is carried out by individual translators, journalists or even terminologists whose training has been linguistic rather than terminological. Therefore, they often need training in the theoretical and methodological basics of terminology science and terminography.

3.9 Summary and Conclusion

In this chapter an overview of LP has been provided and a number of recent trends and some critical approaches and contributions on LP as well as the relationship between LP and TP have been explored with the aim of locating TP in the LP framework. A number of critical models and approaches to LP and TP have been briefly reviewed. The chapter has also described different patterns put forward by Arabic scholars in order to deal with the flood of new terms being created almost every day. Among many other foci, the main focus here was on the major methods of coining new terms (such as derivation, loan translation, compounding, Arabisation and abbreviation). Each method has been examined and defined separately and examples have been given. The chapter concluded with a broad review of the terminological situation in the
ANAC region so as to demonstrate how terminology work is responding to specific challenges raised by the transfer of modern scientific and technical knowledge.

It could be concluded that the issue of terminology in Arabic is compounded due to the lack of terminology standardization, which is due mainly to the nature of Arabic itself. The argument here is not against Arabic (MSA) as a linguistic system. As a matter of fact, Arabic is able to meet the demands for technical texts and computer sciences. According to Raddawi and Rifai (2), “Arabic language is distinguished for its flexible qualities such as synonymy, derivation and analogy. It is able to expand to include the human civilization as well as the technical and scientific knowledge to catch up with the development of science and accept what is new.” The obvious reason for the current dilemma is that the theoretical framework employed by Arabic specialists in conceptualizing the objectives of language and TP and the evaluation of the resulting products (terms) is inadequate. Terminology standardization, therefore, is necessary for the Arabic language in order to develop its functionality in scientific and technical domains and to avoid synonyms and ambiguous communication. This inadequate Arabic theoretical framework for terminology as well as the lack of standardization will be discussed in greater depth in Chapter Six.

It is obvious from the discussion above that the act of term formation is guided by certain strategies. Among these strategies, derivation seems to be the most efficient tool. Its efficiency resides in the fact that it is an effective way to distinguish genuine Arabic words from assimilated ones. Thus, it neither endangers any Arabic existing features, whether syntactical, phonological or grammatical, nor does it risk any aesthetic values of the Arabic language. The use of derivation, therefore, is more appropriate and much needed. However, the task of building
scientific and technical vocabulary is too large to be handled by using only term formation patterns based on the root-pattern system.
Chapter Four

Theoretical Framework and Research Methodology
4.1 Introduction

This chapter is divided into two main parts. The first part gives a comprehensive perspective of Arabic corpus development to users who want to work in this area. The first section of this part gives a brief definition of corpus linguistics and its types. Section two gives an overview of the importance of corpora in terminology studies. Section three provides an overview of some existing Arabic corpora produced by different institutions and companies and also cites some of the common problems researchers face when developing a corpus for the Arabic language from written text. This section demonstrates how the Arabic language lacks textual resources such as corpora and tools for corpus analysis. It also stresses the real need for the development of corpus analysis tools dedicated to Arabic language processing. Section four briefly gives background information on the Word Smith corpus tool (WST) used for analyzing the two study corpora and describes its main functionality.

The second main part of this chapter gives a general overview of the methodology adopted in creating two monolingual Arabic corpora of the special field of computers and the Internet with the aim of studying terminological variation in the Arabic language. This part describes the methodology used for designing the corpora. It presents a detailed account of the structure of the materials and collection procedures. This part also presents theoretical issues such as the questions related to size and representativeness, as well as practical issues related to data collection and legal questions. The final section summarizes and concludes the chapter.
4.2 Corpus Development and Terminology Extraction

4.2.1 General

Corpus linguistics involves the collection of texts that have been gathered in electronic form according to a specific set of criteria (Bowker and Pearson 9). It is not a research area in itself, but rather a processing methodology that can be used for research in all aspects of language, such as assessment, terminology, translation norms, style, and pragmatics. In other words, corpus linguistics is a tool for providing empirical data to support or reject theoretical hypotheses. It can be used to study different aspects of language, such as lexicography, language learning, social language studies, historical linguistics, computational linguistics, terminology, translation and technical writing.

4.2.2 Types of Corpora

Mona Baker (1995) describes different types of electronic corpora that are of specific interest to translation scholars. She distinguishes two different types of corpora: parallel corpora and comparable corpora. A parallel corpus (also called a ‘translation corpus’ among scholars in translation studies) is defined by Olohan (24) as a set of texts in one language and their translations in another language. A parallel corpus can be unidirectional, i.e., STs in language A and TTs in language B, or bidirectional, i.e., STs in language A and translations in language B, and STs in language B and their translations in language A. It can also be bilingual, i.e., STs in language A and their translations in language B, or multilingual, i.e. STs in language A and their translations in languages B, C, D, etc. A parallel corpus in translation studies can be useful in cross-language terminology retrieval and exploration, analysis of translation shifts and
translation strategies, contrastive linguistics, and so forth. Researchers in translation studies can use multilingual corpora in gaining insight as to whether particular translation phenomena are language-specific (idiosyncratic) or widespread across several languages. It is difficult, however, to produce this kind of corpus, since material is seldom translated between multiple languages in equal quantities.

Research using parallel corpora focuses on contrastive linguistic analysis. Much of this research, however, fails to recognize that a translator’s choices may be motivated by something other than language system conventions. It does not take into account, as Olohan (ibid) puts it, any aspect of the translation situation, and it does not allow for the possibility that the translation process may have an effect on language production in translation.

The second type of corpus distinguished by Mona Baker is the **comparable corpus**. The first corpus of this kind was designed by Baker, who defines comparable corpora as “two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages” (234). She states that “we need to study how texts produced in relative freedom (NON-TRANSLATED TL TEXTS) […] differ from texts produced under the normal conditions which pertain in translation (TRANSLATED TL TEXTS)”. First, however, we need to make sure that the translated and non-translated texts are indeed “comparable” in terms of text topic, genre, source type, gender and age of author and translator, etc. Moreover, Baker (ibid) stresses that the texts making up a comparable corpus should be as similar as possible in terms of domain, variety of language, time span and length.
For research using this kind or corpus, Baker suggests capturing patterns that are either related to a specific linguistic feature in a specific language or that are assumed to reflect the nature of translated texts in general and the nature of the process of translation itself (what she terms ‘universal features of translation’). Examples of these features include explicitation—i.e., translations tend to be more explicit on certain levels \(^{18}\) than non-translations—and simplification—i.e., translation content or form (lexico-grammar and/or message) may be more simplified than in non-translated texts—and normalization—i.e., translations unconsciously tend to use language in a more conventional and normal way than non-translations. One may say that besides all these features, translators may consciously strive to make their translations more explicit and more simplified. According to Pápai (145), simplification and explicitation in translation is driven by the translator's conscious or subconscious attempt to respond to the expectations of target readers. Nevertheless, comparable corpora may also be used to investigate aspects of language that are not the result of deliberate practice and of which translators may not be aware.

Although the use of corpora in linguistic studies has a long tradition (e.g., Quirk, Svartvik, Greenbaum and Sinclair 1966), the processing of texts using software systems is relatively new. According to Ahmad and Rogers (724), the first attempt to automate the capture of data from large collections of electronic texts can be dated to the late 1980s. The advantage offered by electronic corpus tools in contemporary linguistics is invaluable. They are used in searching, sorting and displaying a vast quantity of data the corpus contains, as well as in locating and counting instances for analysis. Most importantly, the data could be used for an

\(^{18}\) A study by Pápai (2004) on explicitation as a universal feature of translation found out that explicitation in translation occurs at several levels, such as lexical level (e.g., lexical repetition), grammatical level (e.g., grammatical parallel structures), syntactic level (e.g., addition of participial forms) and textual level (e.g., lexical explanation).
almost unlimited number of studies and analyses. Corpus use is now increasingly understood as a computerized and electronic process in which storing as well as processing of texts is performed automatically or semi-automatically by software systems.

4.2.3 Corpus Studies and Terminology

Terminology studies, as defined by Sager (1990), is the study of and the field of activity concerned with the collection, description, processing and presentation of terms, i.e., lexical items belonging to specialized areas of usage in one or more languages. The use of corpora in terminology studies is relatively new (Wright and Budin 728) as it dates back to the 1980s, especially in languages that use non-Latin scripts, such as Arabic. In the discipline of Arabic corpus linguistics, little research has involved terminology. This can be attributed to the fact that terminology as a science is also a new field of investigation and is still expanding its theoretical foundations and the scope of its application, so Arabic academics and professionals have shown limited interest in terminology as a field of activity distinct from linguistics and later from translation. There has even been considerable reluctance toward terminology work on the part of translators. Among other reasons, this can be attributed to the fact that constructing a terminology database in MSA is time-consuming in comparison to constructing a translation memory (TM), especially with the ongoing resistance to terminology management tools by Arabic organizations and individuals alike. Part of the problem is the Arabic language itself. Not all software applications support Arabic text, and great care must be taken when using Arabic in complex layouts.

In recent decades there has been a trend towards the use of specialized corpus-based terminology research, including terminology management, terminology identification, and
terminology standardization both horizontally and vertically\(^{19}\) (see for instance, Meyer and Mackintosh 1996 and Pearson and Hughes 1988). Sager (762) confirms that “terminology information is no longer extracted from previous lists or by individual searches, but rather from a body of relevant textual materials called a corpus.” This has changed the nature of the tasks terminologists need to undertake. Therefore, the nature of basic terminological problems, such as the process of identifying terms, has changed, and thus the attitude of people towards terminology has also changed. Today, this process tends to be automated and is referred to as automatic term extraction and sometimes semi-automatic term extraction as it requires some editing and processing (see Ahmad and Rogers, *Corpus Linguistics and Terminology Extraction* 2001).

The relevance of corpus-based studies for terminology, according to Ahmad and Rogers, “can be simply illustrated by reference to neologisms and archaisms (quantifiable through frequency measures of new terms and obsolete terms respectively) and compound terms, the staple of any terminology enterprise (accessible, for instance, through concordance searches, statistical tests of collocational patterns, or the identification of typical patterns characterizing possible forms)” (730). Bowker and Pearson (2002) also state that corpora are useful in studying LSP neologisms because corpora are kept up to date more easily than printed resources. They note, however, that only neologisms that have been formed by merging words into new combinations can be identified with automatic tools; other neologisms formed by identifying a

\(^{19}\) Horizontal standardization is defined as the standardizing of procedures or principles of terminology work, whereas vertical standardization is the standardizing of terminology for a specific subject field. Horizontal standardization is established by experts in terminology science, including linguists, language planners, terminologists, professional translators, and so forth, whereas vertical standardization is the product of subject specialists of a given field.
new meaning must be manually identified. This is the main factor that makes fully automated terms extraction an unattainable dream.

Many researchers (Bowker and Pearson (2002), Sinclair (1991), Ahmed and Rogers (1992), etc.) discuss how monolingual specialized corpora can be used to identify contextual examples and information on terminological usage. Ahmed and Rogers highlight the fact that “Special-purpose corpora can provide evidence not only on linguistic variation for the translator, whose work is text-based, but also for the terminologist, whose objectives may be normative. If a corpus is well-designed, then the user will be able to select both the range of texts to be processed and the method of processing. For purposes of standardization, for instance, a frequency count of synonyms can provide useful distributional evidence indicating statistically preferred terms.” (36)

The advantages of specialized corpora are discussed by Bowker and Pearson (2002). They point out that corpora are deemed to overcome many of the shortcomings of other resources commonly consulted by translators to find equivalents, such as dictionaries, because they are an extensive source of authentic usage information and are easy to update and to search. The particular usefulness of specialized corpora is that they show how the environment surrounding the term can give descriptions and/or definitions of the concept to which it refers. The corpus can also be useful in gaining a broader understanding of the concepts themselves.

Obviously, the optimal design of a corpus is largely dependent on the purpose for which it is intended (Kennedy, 70), but there are general factors that need to be taken into account when building specialized corpora. Pearson (1998, cited in João Cotter 500) defines several relevant factors such as size, text type and origin, authorship, factuality, technicality, audience, intended
outcome, setting and topic. Bowker (596) also lists numerous factors that must be taken into consideration when collecting and compiling specialized corpora: 1. General vs. idiosyncratic usage (i.e., whether the term in the corpus is a generally accepted term in the field or whether it is used by only one particular author). 2. Life cycle of the term (Sager 59 maintains that most terms are initially provisional until they become widely accepted and so incorporated into the lexicon of a particular language). 3. The purpose of the text (i.e., whether the text has been written for an expert audience or for a more general language audience). 4. External alternation to the text (i.e., whether the text has been subjected to external alteration after it left the original author’s hands).

4.3 Arabic Corpora

There are some corpora for Arabic that have been developed for special purposes (Hoogland 1996, Zemanek 2001, De Roeck 2002, Maamouri & Cieri 2002, Buckwalter 1986, Izwaini 2003). These corpora vary in size and the genres they include. However, they are all limited in domain and in coverage of the geographical variation in MSA (Al-Sulaiti and Atwell 1). One of the first modern electronically readable corpora is the ‘Buckwalter Arabic Corpus’ developed by Tim Buckwalter. Collecting texts for this corpus started in 1986 before electronic Arabic texts appeared on the web. The corpus consists of about three million words transcribed from public Arabic newspapers and magazines such as Al-Ahram (Cairo), Al-Hayat (London), Al-Bayan (UAE), Al-Nahar (Beirut) and Al-Raya (Qater). Another Arabic corpus available for public use is the Qur’anic Arabic Corpus developed by the Arabic Language Computing Research Group at Leeds University. This corpus is an annotated linguistic resource consisting of 77,340 words of the Arabic Qur’an. It provides three levels of analysis: morphological
annotation where each word of the Qur’an is tagged with its part of speech (POS) as well as multiple morphological features; syntactical Treebank, which shows the Arabic grammar (إعراب) ‘i’rab’, syntax and morphology; and semantic analogy. The following screenshot is an example that shows syntactical Treebank in the Qur’anic Arabic Corpus:

![Syntactical Treebank Example](image)

Figure 4.1: Qur’anic Arabic Corpus syntactical Treebank

There are some other Arabic corpora, mainly in the form of raw data, which have been produced by different institutions and companies for different purposes. An exhaustive description of all of them would be beyond the scope of this study. In the following table, Al-Sulaiti and Atwell (2001) briefly list the main written Arabic corpora that have been developed and are available online for research purposes, both free of charge and fee-based:
<table>
<thead>
<tr>
<th>Name of Corpus</th>
<th>Source</th>
<th>Size</th>
<th>Availability</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buckwalter Arabic Corpus 1986-2003</td>
<td>Tim Buckwalter</td>
<td>2.5 to 3 billion words</td>
<td>Registration fee is required</td>
<td>Public resources on the web</td>
</tr>
<tr>
<td>Al-Hayat Arabic data set</td>
<td>ELRA (European Language Resources Association</td>
<td>18,639,264 million words</td>
<td>Available on ELRA’s catalogue</td>
<td>Articles on cars, computer, sports, news, economics and science</td>
</tr>
<tr>
<td>An-Nahar text corpus</td>
<td>ELRA (European Language Resources Association</td>
<td>24 million words</td>
<td>Available on ELRA’s catalogue</td>
<td>Arabic articles in (Lebanon) from 1995 to 2000</td>
</tr>
<tr>
<td>Nijmegen Corpus</td>
<td>Nijmegen University</td>
<td>2 million words</td>
<td>Free</td>
<td>Sources such as Al-Wasat and Al-Arabi magazines, some fiction and non-fiction sources</td>
</tr>
<tr>
<td>CLARA (Corpus Linguae Arabicae)</td>
<td>Institute of Ancient Near Eastern Studies at Charles University,</td>
<td>37 million words</td>
<td></td>
<td>Periodicals, books and from other miscellaneous written materials.</td>
</tr>
<tr>
<td>Egypt</td>
<td>Centre for Language and Speech Processing at John Hopkins University</td>
<td>Unknown</td>
<td>Available free on the web</td>
<td>Parallel texts of the Qur’an in English and Arabic</td>
</tr>
<tr>
<td>DIINAR Corpus</td>
<td>DIINAR-MBC (DLictionnaire INformatise de l’ARabe, Multilingue et Base sur Corpus</td>
<td>10 million words</td>
<td>Non-free</td>
<td>Multilingual lexical database in Arabic, English and French using high-level multilingual language resources</td>
</tr>
<tr>
<td>English/Arabic Parallel Corpus</td>
<td>University of Kuwait</td>
<td>3 million words</td>
<td>Available in CD-ROM and on the Web, but restricted to users in the University of Kuwait.</td>
<td>A’laam Al-Ma’rifa’ (the World of Knowledge) published by the National Council of Culture, Arts and Letters in Kuwait (NCCAL)</td>
</tr>
<tr>
<td>SOTETEL Corpus</td>
<td>Société Tunisienne d’Entreprises des Télécommunications</td>
<td>8 million words</td>
<td>Registration fee is required</td>
<td>Variety of genres such as literature, academic and journalistic writings</td>
</tr>
<tr>
<td>Classical Arabic Corpus (CAC)</td>
<td>University of Manchester Institute of Science and Technology</td>
<td>5 million words</td>
<td></td>
<td>Short poems from the period of the advent of Islam up to the end of the eleventh century</td>
</tr>
<tr>
<td>General Scientific Arabic Corpus (GSAC)</td>
<td>University of Manchester Institute of Science and Technology</td>
<td>10.7M words</td>
<td></td>
<td>Kuwaiti magazine site ‘Science and Technology’</td>
</tr>
</tbody>
</table>

Table 4.1: Arabic corpora

4.3.1 Problems in Corpus-building for Arabic

The Arabic language is the fifth most spoken language in the world, after Chinese (Mandarin), Spanish, English and Hindu-Urdu; it is spoken by approximately 250 million people.

20 ELRA (European Language Resources Association) is an organization located in Paris from which corpora can be purchased for academic or industrial purposes. (See http://www.elra.info/).
and is one of the six official languages of the United Nations. Nevertheless, its language resources are somewhat deficient. Few readily available language resources such as corpus analysis tools function completely and correctly with Arabic. This can be attributed first to the fact that research priorities within languages, particularly corpus linguistics, have focused on languages with Latin script, mainly European languages. Thus, most linguistic tools used to be restricted to American Standard Code for Information Interchange (ASCII) character sets (i.e., Latin alphabets). However, in recent years, there has been a shift towards developing tools in languages such as Chinese, Japanese, Korean and Arabic.

Second, Arabic script is considerably more difficult to process than European languages such as English and French. In fact, the difficulties that the Arabic language pose to researchers are not limited to its sociolinguistic status or ‘triglossia’ (described in detail in Chapter Two), but also to its inherent linguistic structure. The following are the commonly perceived difficulties in processing Arabic.

4.3.1.1 Bi-directional Nature of Arabic Script

Arabic is written from right to left (RTL), which is in acute distinction to the majority of other languages, written from left to right (LTR). However, the bi-directional feature, also referred to as (BiDi) (a mixture of right-to-left and left-to-right direction), is also required when processing Arabic script—for instance, when adding Hindi numerals, which are currently used in most countries in the AW. Table 4.2 shows Hindi numerals and their equivalents in Arabic numerals, which are used in most European languages. Unfortunately, not all language resources have truly reliable right-to-left and left- to-right combined support, and great care must be taken when using Arabic in complex layouts for entering, editing and displaying text. Some
applications may not have any of the support that is required for BiDi orientation and treat Arabic text as if it ran just from LTR or RTL, so they garble the Arabic text. Others claim to support this orientation, but do not reliably switch back and forth appropriately, so text often ends up backwards or in mirror image (see Figure 4.2). However, there is a recent general trend to implement BiDi text components in most modern operating systems and platforms such as Java.

<table>
<thead>
<tr>
<th>Hindi numerals</th>
<th>۰</th>
<th>۱</th>
<th>۲</th>
<th>۳</th>
<th>۴</th>
<th>۵</th>
<th>۶</th>
<th>۷</th>
<th>۸</th>
<th>۹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic numerals</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 4.2: Hindi and Arabic numerals

Compiling an Arabic corpus by means of a corpus tool that does not meet the requirement of BiDi Arabic text components would result in a corrupted text. The Arabic text would appear with spaces between letters of the same word, which prevent the texts from being viewed and thus read properly. The following screen shot shows the problem:

Figure 4.2: Distortions in Arabic text
Many corpus tools have in fact been updated to work properly with the BiDi feature of Arabic text, but still require additional configuration, such as enabling right-to-left support at the level of the operating system being used.

### 4.3.1.2 Peculiar Features of Arabic Script

Arabic characters are represented differently and take different shapes depending on whether they occur at initial, medial or final position in a word. For example, the character ‘ْه’ ‘hah’ may have four different basic shapes. It has an initial shape (ٍْ)، a medial shape (ٌْ)，a final attached shape (ٌْ) and a final detached shape (َْ).

<table>
<thead>
<tr>
<th>Representative shape in code chart</th>
<th>Possible shapes in context</th>
</tr>
</thead>
<tbody>
<tr>
<td>ٍْ</td>
<td>Final (detached)</td>
</tr>
<tr>
<td>َْ</td>
<td></td>
</tr>
<tr>
<td>ُْ</td>
<td></td>
</tr>
<tr>
<td>ٌْ</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Different shapes of Arabic character ‘hah’

Table 4.4 gives examples of the four shapes and positions of the Arabic character ْ ‘hah’ (indicated in red) in Arabic words:

---

The screenshot below shows the same Arabic texts, but properly displayed:
All Arabic language processing systems implement automatic selection of the correct shape according to its position in the word, so that the user does not have to select the correct shape manually. Moreover, some Arabic letters have the same shape and can only be distinguished by adding certain marks such as dots, as in (ب, ت, ث, ج, خ), or a hamza21، as in (١، ١، ١، ١) and (ب، ى). When using corpus software, such characters have to be written exactly the same way in the concordance search window and/or KWIC display so that they can be found. To handle this problem in Arabic language processing systems, researchers have implemented a normalization strategy. For instance, Larkey and Connell (2001, cited in Farghaly and Shaalan 14) replace the initial ‘الف’ with a ‘حمزة’ above or below (١ and ١) with simply an ‘الف’ (١). They also normalize the ‘الف مدا’ (١) with a ‘بَاَلْف’ (١). However, Farghaly and Shaalan (ibid) argue that although normalization improves recognition by solving the variations in Arabic script, it increases the probability of ambiguity. For example, if we normalize the ‘الف’ in (لاَن), it becomes hard to distinguish between (لاَن) and (لاَن). While the first is translated as ‘to be flexible,’ the second could be translated as because or since. Moreover, the same word written with or without the hamza can be recognized by the concordance program as two different words.

21 Hamza in Arabic orthography is a sign used together with or independent from (١) alif. When it is used with alif its purpose is to represent a glottal-stop sound. It may appear on the alif in certain circumstances to indicate that the hamza is one of the 28 Arabic alphabets independent from the alif.
4.3.1.3 Arabic Morphological Complexity

Although there are fewer differences between Arabic and English morphological systems, Arabic morphology is substantially more complex than English. In Arabic, as in English, adding an affix to a root can produce a new word. Derivational morphemes form new words either by changing the meaning of the root to which they are attached, e.g., *able* and *disable*, or by changing the word class that a root belongs to, e.g., *certain* and *certainly*. In Arabic almost all words are derived from roots by the addition of suffixes, prefixes and infixes. For instance, from the root ‘كتب’, denoting *writing*, we could extract the following verbs:

<table>
<thead>
<tr>
<th>Arabic word</th>
<th>POS</th>
<th>Transliteration</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>كتاب</td>
<td>verb</td>
<td>kataba</td>
<td>he wrote</td>
</tr>
<tr>
<td>يكتب</td>
<td>verb</td>
<td>yaktub</td>
<td>he writes</td>
</tr>
<tr>
<td>يكتبون</td>
<td>verb</td>
<td>yaktubon</td>
<td>they write</td>
</tr>
<tr>
<td>كتبوا</td>
<td>verb</td>
<td>katabu</td>
<td>they wrote</td>
</tr>
<tr>
<td>يكتبان</td>
<td>verb</td>
<td>yaktuban</td>
<td>they write (dual)</td>
</tr>
<tr>
<td>كتاب</td>
<td>verb</td>
<td>kataba</td>
<td>they wrote (dual)</td>
</tr>
<tr>
<td>نكتب</td>
<td>verb</td>
<td>naktubu</td>
<td>we write</td>
</tr>
<tr>
<td>كتبنا</td>
<td>verb</td>
<td>katabbna</td>
<td>we wrote</td>
</tr>
</tbody>
</table>

Table 4.5: Arabic derivation (verbs)

Many nouns can also be extracted from this root, such as the following:

22 In Arabic, there are three grammatical numbers, singular, dual and plural. The dual indicates two things or two persons. It is usually formed by adding certain suffixes to the singular noun, e.g., يدُ يدٌ يدٍ a hand and يدانُ يدانٍ يدانٌ two hands. The dual in this example is formed by adding the suffix [ٍ] to the singular [ُ] hand.
Corpus analysis on European languages is often performed on parsing out roots or lemmas in order to identify actual usage despite textual changes involving affixation. In Arabic corpus, however, performing a concordance search for a root like ‘ كتاب’ would return too many matches, and it would clearly be very frustrating to have to go through all these matches, many of which are not associated with the ‘كتاب’ root, simply because they happen to contain these three consonants (Roberts and Al-Sultiati 41). One relative solution for this problem is ‘root extraction’, a method for detecting roots from which derivations are made by basically removing the suffixes and prefixes attached to the given word and reducing it to its bare form. However, this would turn the concordance search into a laborious and time-consuming process, as every single lexical unit has to be checked and manually manipulated.

### 4.3.1.4 Arabic Vocalization

The phenomenon of *vocalization* (i.e., texts with or without Arabic diacritics ‘تشكيل’ taṣkil) is difficult to visualize in English and other languages that use Latin characters. In English, for instance, vowels are part of the alphabet. Vowel fonts are stable whenever they occur and are used consistently in combination with consonants. Letters can represent either spelling patterns

<table>
<thead>
<tr>
<th>Arabic word</th>
<th>POS</th>
<th>Transliteration</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>كتاب</td>
<td>Noun</td>
<td><em>katib</em></td>
<td>writer</td>
</tr>
<tr>
<td>مكتب</td>
<td>Noun</td>
<td><em>maktab</em></td>
<td>office</td>
</tr>
<tr>
<td>كتاب</td>
<td>Noun</td>
<td><em>kitaab</em></td>
<td>book</td>
</tr>
<tr>
<td>مكتوب</td>
<td>Noun</td>
<td><em>maktoob</em></td>
<td>letter</td>
</tr>
<tr>
<td>مكتبة</td>
<td>Noun</td>
<td><em>maktaba</em></td>
<td>library</td>
</tr>
<tr>
<td>كتاب</td>
<td>noun</td>
<td><em>kutāb</em></td>
<td>booklet</td>
</tr>
</tbody>
</table>

Table 4.6: Arabic derivation (nouns)
or markers. Spelling patterns are the letters that are pronounced, whereas markers are not pronounced and do not have a direct relation to the pronunciation of the word, although sometimes they signal sounds. A common example is the letter ‘e’ at the end of the word; it is very often silent, but it usually signals a change in the pronunciation of the preceding vowels in the word so that they are pronounced as long vowels (diphthongs).

Arabic, on the other hand, has a characteristic that could be of potential interest to translation studies research, as translators may encounter comprehension and translation difficulties when they deal with Arabic texts that are not presented with vocalizations, because without vowels “reading in Arabic orthography is a difficult mission even for skilled readers as many words are visually and orthographically homographic; they look the same, but they carry different meanings” (Abu-Rabia 631).

Arabic has a system of supplementary vowel symbols different from English and any other European language. In other words, Arabic has 28 letters that are all consonants. Some of them (א, ו waw, ي ya), however, can be considered as long vowels in some cases, and they are part of the Arabic alphabet. Unlike English, short vowels in Arabic are not part of the alphabet. They are specific signs تشكيل taṣkīl added to take the form of tiny marks placed either above or beneath the (consonant) letters. In Arabic, there are three main short vowel signs (א, א, א) and other signs حركات harakat (movements) that indicate the silent sounds, the prolongation and the double pronunciation of a consonant. Consider the following table.
<table>
<thead>
<tr>
<th>Sign</th>
<th>Name</th>
<th>Sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>ﻝ ﻝ</td>
<td>Fatha</td>
<td>A</td>
</tr>
<tr>
<td>ﺚ ﺚ</td>
<td>Kasra</td>
<td>I</td>
</tr>
<tr>
<td>ﺞ ﺞ</td>
<td>dhamma</td>
<td>U</td>
</tr>
<tr>
<td>ﺦ ﺦ</td>
<td>sadda</td>
<td>Doubles the letter underneath</td>
</tr>
<tr>
<td>ﻧ ﻧ</td>
<td>sukoon</td>
<td>Silent sound</td>
</tr>
<tr>
<td>ﺧ ﺧ</td>
<td>madda</td>
<td>Prolonging the (a) sound</td>
</tr>
</tbody>
</table>

Table 4.7: Summary of Arabic short vowel signs

In Arabic orthography, vowels are optional, which means texts are introduced with or without short vowels. A skilled Arabic reader reads both voweled and unvoweled texts fluently. For example, the Arabic word for university is جامعة ﻪﺎم’ ﺍه, which can be introduced to the Arabic reader as جامعة ﺍه without vowels or جامعة ﻪ ﺍه with vowels. As can be seen, there is a little stroke fatha above the first character, which indicates a short (a) sound, another little stroke kasra beneath the third character, which indicates the (i) sound, a small circle sukoon above the fourth character that indicates there is no short vowel at this point, and waw-like sign dhamma above the final character, which indicates the (u) sound.

Unvowled Arabic is becoming the norm in MSA writings, especially in electronic texts, including newspapers, magazines and textbooks. The fully vowelized orthography is used only in classical Arabic scripts such as the Qur’an and certain formal texts such as poetry. Vowelized Arabic texts “are easy to phonologically decode since the letters and diacritics have highly consistent and reliable grapheme-to-phoneme (letter to sound) correspondences” (Fender 102). For example, the trilateral radicals ﻪ ﻪ ﻪ can mean science or knowledge if pronounced as ﻪ ﻪ ﻪ elim, a flag if pronounced as ﻪ ﻪ ﻪ َ ﻪ alam, known if pronounced as ﻪ ﻪ ﻪ َ ulim, to be taught if pronounced...
as ٌعَمَّ َullim, etc. On the other hand, the orthographic ambiguity causes problems in the tokenization process\textsuperscript{23} used to determine frequency, because these four words would be counted as one, resulting in unreliable statistics for the frequency of such words. This is also one of the problems that are very difficult to solve. Therefore, any morphological analyzer for Arabic should be able to recognize and disambiguate unvowelized text.

Many tools and corpora have been developed to handle Arabic texts without vowels, but any attempt to apply these tools to vowelized Arabic texts would be far from accurate, since there is significant ambiguity between unvowelized words that have the same lexical representation. For example, the following table shows the 10 top words and their frequency in Tim Buckwalter’s Arabic corpus based on three separate counts (February 1999, February and August 2002). Although some word forms are unambiguous and easily merged (e.g., إلى and إلى are the same), most word forms require contextual analysis to be disambiguated because of the absence of vowels (e.g., مَن and مَن; while the first is a preposition meaning \textit{from}, the second is an interrogative pronoun meaning \textit{who}).

\textsuperscript{23} The process of breaking up the corpus into list of words to find frequency occurrences of terms.
Another problematic feature that makes Arabic frequency statistics unreliable is proper names, which cannot be distinguished with a capital letter, since capitalization does not exist in Arabic, unlike Latin script-based languages where sentences begin with capital letters and end with a period. Consequently, it is not possible to identify proper names and words from other languages by means of a corpus wordlist where words are listed independently of their immediate context. The absence of capitalization also makes Arabic name entity recognition more difficult. For example, the word ‘زَهْرَة’ zahra means *Venus*, a girl’s name, and it also means a *flower*. The average reader is able to disambiguate these meanings based on the local context in which they are used, but automated recognition systems cannot ‘intuit’ context.

### 4.3.1.6 Punctuation

Languages have different punctuation patterns. Due to the absence of strict punctuation rules in Arabic, recognizing sentence boundaries in a running text by applications such as corpus
tools, machine translation, translation memory and terminology retrieval is a more difficult task in Arabic than it is in languages like English. In fact, conjunctions such as (و) waw and (ف) fa\(^{24}\) can combine not only several words as in English, but also sentences and sometimes paragraphs. Consider the following example:

A new and modern Arab regional order is emerging along with a changing form of speech devoid of the previous demagogy. This change was confirmed by the “Riyadh Declaration,” presented at the end of the Arab Summit in March 2007, which laid, for the first time, the foundation of an “Arab basis” for the culture of integration versus the culture of division which still prevails in the region. It also ascertained that “Arabism is not a racial idea but a cultural one based on spiritual,

\(^{24}\) ‘fa’ ف is a connective particle, which can be translated as ‘and’, ‘so’ or ‘then.’
moral and humanitarian values, enriched by diversity, pluralism, and openness to other human cultures alongside rapid advances in education and technology...

Unlike *and* in English, the equivalent Arabic (و) *waw* is not considered as a separate unit by many corpus analysis tools, including WordSmith Corpus Tool WST (a detailed description of this tool will be introduced in the following section). It is generally considered as a prefix attached to Arabic words. Consider the following screenshot:

Figure 4.4: (و) *and* in Arabic corpora

WST was unable to identify that the conjunction و *waw* *and* attached to words in the above screenshot as separate units. Only the و *and* *wassi’a wide* is a part of the stem of the word.

The same thing can be said about the Arabic definite article ال *the*, some conjunctions and coordinators and pronouns. They are all normally attached to other words such as nouns, verbs and prepositions, so they are not considered by corpus tools as separate units.

121
Before concluding this section, it is worth noting that the common problems cited here comprise only some of the issues that corpus users face in developing a corpus for Arabic language from written text. There are many other issues that are not related to the Arabic language system and whose source of difficulty is not just the Arabic script itself. For instance, Arabic is an underdeveloped language with regard to computational linguistic research. There has been insufficient linguistic research in Arabic to create computer resources needed in a modern computing environment (Yassin 1). There are no grammar checkers for Arabic, no optical character recognition (OCR), and, most importantly, no powerful linguistically-aware search engines or string-processing utilities to handle Arabic. As a result, “Arabic poses some of the greatest web localization challenges because of poor software support” (Yunker, 399). Part of the problem is the traditional Arabic grammar widely employed today. MSA grammar and style are rather complex compared with other languages such as English and French (see Chapter Two).

4.4 Wordsmith Corpus Tool WST

The corpora for this study were managed using Wordsmith Tools corpus analysis software, an integrated suite of programs devised by Mike Scott (Scott 1996) for analyzing how words behave in texts. Nowadays, there are many corpus tools available to carry out research on text behavior—for example, ParaConc, AntConc and MonoConc. Most of them offer similar applications and the possible functions expected of any corpus software program. The differences among them are mainly in the way data are displayed, their ability to perform specific functions such as concordance search, and most importantly, their ‘friendliness’ towards languages with non-Latin scripts, such as Arabic, Chinese and Persian. WST is more popular as
monolingual software for lexical analysis than other corpus tools. It handles a good range of mostly European and non-European languages. In addition, WST is easy to use and thus a convenient choice, especially for less experienced users. It has been widely used to produce specialized corpora for technical term extraction. WST allows three basic operations to be performed on selected texts, as described below.

4.4.1 Word Frequency List

A word frequency list (also referred to as a frequency table) displays the words occurring in a corpus along with the number of times each word appears. By means of word frequency, it is possible to analyze term variation and term formation from a single root. The following screenshot is an example of word frequency list representation in WST:

![Wordlist as displayed by WST](Image)

Figure 4.5: Wordlist as displayed by WST
However, a word frequency list is often a list of frequent words in general language, which a researcher does not want to include in his/her analysis (for example, common function words like the, of, is, for, etc.). Terminology and LSP researchers usually look for the frequency of terms and other words that are specialized in a special area of discourse. WST’s ‘stop list’ feature can exclude these function words and other common words from the frequency wordlist (for more details about this function, see Scott 2004.) Moreover, term frequency does not always imply that the term is relevant in the discourse; it may simply be due to the author’s idiosyncrasies, as well as her/his biases towards certain terms. With the WST word list function, which indicates the frequency of each term in every text, it is possible to check whether the term is frequent only in a given text from the corpora or frequent in different texts.

4.4.2 Concordance

A concordance “is a list of all the examples of the target item (the linguistic phenomenon being searched for), normally accompanied by enough context to enable a human being to study the item’s occurrence in detail” (Leech and Fligelstone, 1992). It indicates whether there is a statistically significant association between two words in a text. The main advantage afforded by the concordancing tool is that it enables users to see terms in a variety of contexts simultaneously, which in turn allows them to detect various kinds of linguistic and conceptual patterns that are sometimes difficult to spot in isolated electronic as well as printed resources (e.g., meanings of terms, related terms, typical phrasal patterns; Bowker 2002). However, developing a concordancing tool that can reliably cope with the Arabic language has proved difficult (Roberts, Al-Sulaiti and Atwell 39). The most significant problem is that the output of Arabic concordances in many tools (e.g., MonoConc) is in incorrect order due to the fact that
these tools do not account for the BiDi nature of Arabic text that was discussed in the previous section.

Figure 4.6: Arabic concordance as displayed by WST

WST seems to be a suitable choice for the purpose of Arabic text analysis. I was able to generate frequency list of words and find relations among selected words using the WST concordance function. The screenshot (Figure 4.6) shows concordance results for the word `نظام` nidham system displayed in concordance lines. The display can be resorted in different ways, e.g., ascending and descending alphabetical order. It was also possible to mark or delete lines and search for multiword terms.

4.4.3 KWIC (Key Words in Context), and the reference to the text (source).

This output makes it possible to derive conclusions with respect to both intra- and inter-regional variations. KWIC output is particularly efficient in that it simultaneously presents multiple
examples of a term and draws multiple examples of terms together in a single representation in order to highlight patterns of terminological variation.

4.5 Research Methodology

4.5.1 General

As discussed in the introductory chapter, this study is based on two methodological approaches that are connected by the attempt to answer the research questions and to address their assumptions. The first is a descriptive analysis based on a careful examination of the ANAC sociolinguistic status (described in Chapter Two) so as to draw a clear picture of that diverse landscape. The purpose was to demonstrate the complexity of the questions involved in terminology policy in the region and to help to explain the social, political, geographical and economic factors that may have an impact on term formation. Most information was taken from published documents and government reports so as to draw a clear picture of the linguistic and terminological status in the region. Due to the lack of references about the general status of terminology work in the AW both at theoretical and practical levels, collection relied on information that is available on the World Wide Web.

The second methodological approach was to investigate the influence of colonialism and dialectology on term formation and use and to detect terminological variations in the context of computers and the Internet. The data for this investigation was collected from two small monolingual Arabic specialized corpora compiled from different websites (specialized in computer and software industry, web hosting and web design and Internet service) and

25 The two corpora used in this study are considered specialized as they deal with the area of computers and Internet, have a special purpose as they are intended to examine terminology variation related to computers and Internet, and are monolingual as all compiled texts are written in the Arabic language.
publications (such as manuals, books, tutorials and journalistic articles) from various countries in the ANAC. In the following section a detailed description of the steps and procedures of the corpora creation is given.

4.5.2 Corpus Design and Compilation

The corpora compiled for this study were two ‘raw’ comparable electronic corpora of original Arabic texts that represent the two foreign-language orientations (French and English, for more details on foreign-language orientation, see Chapter Two). Raw corpora means that the texts themselves are compiled ad hoc without any kind of tags or markup\(^{26}\). This kind of corpora should be distinguished from annotated corpora where interpretive linguistic information has been added in order to extend the range of research questions that the corpora may address\(^{27}\). Annotated corpora are also distinct from corpus markup corpora, which provide relatively objectively variable information regarding the components of the corpora and the textual structure of each text (for more information about the three types, see in particular Al-Sulaiti and Atwell 2004.)

The first corpus represents the French-language orientation in Morocco, Algeria, Tunisia and Mauritania, and the second represents the English-language orientation that exists in Libya, Egypt and almost all of the rest of the AW. The corpora were based on authoritative documents such as research articles, technical reports and specialized journals and magazines. All the texts

\(^{26}\) This decision was taken in accordance with the objective of the study as it was not necessary to tag the corpora for parts of speech.
in the corpora were freely downloadable from the Web\textsuperscript{28}. The web sites comprising the two created corpora are listed in Appendix I.

The text selection process was driven mainly by the amount of data available. This task was expected to pose a number of challenges and difficulties in terms of quality and quantity due to the severe lack of relevant electronic Arabic resources on the WWW in general. The issue is even more serious when it comes to finding specialized Arabic texts. The first and most laborious task, therefore, was to find a representative quantity of quality MSA texts that deal with computer and Internet contexts available in electronic format over a wide geographical area to fulfill the purposes of this study. In order to obtain a large amount of representative data, texts were mostly selected from a variety of information technology (IT) sources, but they also include non-specialized documents such as news reports for governments, journal articles, and newspaper articles available on the WWW.

The following are the main features of the corpora:

- To limit confounding variables, the two representative corpora are intended to be balanced and comparable in terms of text topic, number of texts and words, date of publication, target readership, etc. Most importantly, the eligible texts cover the same subject field (computers and the Internet) from different sources in all the ANAC countries. Table 4.9 below details the most relevant statistical information generated by WST concerning the composition of the two corpora, i.e., corpus size, the member of token and types, the type token ratio TTR and number of sentences. The table shows that the two corpora are comparable in terms of total

\textsuperscript{28} Several Arabic corpora have been built from written texts compiled freely from the Internet such as the Buckwalter Arabic Corpus BAC (1986-2003), Leuven Corpus LC (1990-2004), CLARA (1997), General Scientific Arabic Corpus GSAC (2004), Classical Arabic Corpus CAC (2004) and Corpus of Contemporary Arabic. CCA.
number of tokens (words). It indicates that C1 has a slightly higher number of tokens than C2. 250,360 tokens were obtained for C1 and used for the word list, whereas 246,624 tokens were obtained for C2 and used for the word list. C1 and C2 were matched on the basis of the total number of words but not on the total number of texts included. In other words, the total number of words per corpus was equaled by including more texts in cases where one corpus has fewer words than the other. This may explain the difference between the two corpora in terms of the total number of texts included (33 texts for C1 and 30 texts for C2). It also justifies the slight difference in numbers of sentences (12,583 sentences in C1 and 11,287 sentences in C2).

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Total no. of tokens</th>
<th>Tokens used for the list</th>
<th>TYPE (distinct words)</th>
<th>TTR%</th>
<th>No. of texts</th>
<th>No. of Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>250,360</td>
<td>250,360</td>
<td>128,369</td>
<td>33.85</td>
<td>33</td>
<td>12,583</td>
</tr>
<tr>
<td>C2</td>
<td>246,624</td>
<td>246,624</td>
<td>130,269</td>
<td>40.45</td>
<td>30</td>
<td>11,287</td>
</tr>
</tbody>
</table>

Table 4.9: Statistical information on C1 and C2

WST provides a measure that helped in determining the level of specialization of texts used in the two corpora, i.e., Type/Token Ratio (TTR). TTR was determined by dividing the total number of types (i.e., different words in the corpus), which is 128,369 for C1 and 130,269 for C2, by the total number of tokens (i.e., total number of words). The TTRs for the two corpora are different: 33.85% for C1 and 40.45% for C2. C1 has a lower TTR than C2, which means that lexical diversity of C1 is higher than that of C2. In other words, C2 is less varied in words than C1, whereas C2 shows a larger amount of lexical variation, but in general both corpora have low TTR. This might be interpreted as an indicator of the high degree of specialization of texts used in the two corpora.
These results might not confirm the previous claim about the dynamic and complex nature of the computer and Internet domain. However, the low TTR might be also attributed to the nature of the Arabic language itself, which utilizes lexical repetition as a major stylistic device to emphasize the message, thus creating an exceptionally high level of explicit cohesion compared with other languages.

- The compiled documents belong to institutions and organizations from different countries in the region, namely: university, company and organization websites, formal online technical support, IT magazines and periodicals, as well as computer and Internet websites and journals. This was performed by using Google “Advanced Search” Regional Function (http://www.google.com29). The search was also widened to consider the ‘Internet site’s country domain’ like ‘.ly’ for Libya, ‘.eg’ for Egypt, ‘.dz’ for Algeria, ‘.tn’ for Tunisia and ‘.ma’ for Morocco. The question that can be raised here, however, is how well Google differentiates the country or region from which a web page originates. Google explains that

We rely largely on the site’s country domain (.ca, .de, etc.). If an international domain (.com, .org, .edu, etc.) has been used, we'll rely on several signals, including IP address, location information on the page, links to the page, and any relevant information from Google Places.

(https://support.google.com/webmasters/answer/62399?hl=en)

The low figures in Table 4.9 above might be the result of the countries’ small Internet presence. According to Ping Dom’s statistics (accessed on November 23, 2013), there were only 276

29 Google Advanced Search offers several ways to narrow and filter search in order to get more precise and useful results. For more details about how to use this function go to (https://support.google.com/websearch/answer/35890).
million web pages in the `.eg` (Egypt) domain and 97 million in the `.ma` (Morocco) domain, compared with Ireland, where the number of web pages in the `.ie` domain was 4,1112 million pages. While the population in Egypt is 83 million people, the population in Ireland, according to an 2013 estimate, is only 4.5 million people.

Another problem that affects data collection is that websites, particularly computer and Internet magazines and periodicals, are frequently updated and materials are available only for a limited time. Therefore, the collection process was carried out within one period of time, August 2012, to reduce the possibility of changing and/ or deleting sources.

From what preceded, it is worth mentioning that although data were collected from all the countries of the ANAC region, I make no claim that these data are representative of all technical discourse in those countries. This is because technical texts are written by different authors, representing various aims and beliefs at different times. As Van Mol (117) puts it, it is almost impossible to compile a corpus that would represent the whole linguistic reality of a speech community.

- Resources are presented on the web in three basic presentation forms: HTML and XML text, and PDF. Documents in PDF, XML and HTML format were located in MS Word by means of copy and paste before being converted into plain text format and saved as Unicode. Using Unicode as a standard code page is crucial here, because it unifies the automation processing of non-Latin languages like Arabic. In the case of HTML and XML format, the procedure for including the material in the corpora is ‘copy and paste,’ and in the case of PDF, it involves downloading and saving the material in text format. This is a critical feature in this

---

30 For the particular purpose of this study Unicode UTF-8 has been used because it is becoming the dominant character encoding for the WWW and reliably expresses and retains orthographic features of MSA.
case for two reasons. First, the WST can index and retrieve Arabic texts only in plain text format (*.txt). Second, Unicode is an encoding scheme that allows the use of the full variety of scripts for human languages, including Arabic, which is not accommodated by earlier encoding schemes. Although converting these texts’ formats into plain text format is quite straightforward, Arabic texts require special handling during the conversion of the texts from the PDF, HTML, DOC or other file formats into the .txt format for making the texts processable by WST. Many changes occurred during the conversion process. Words, and in some cases paragraphs, can switch their positions during the conversion process. As a result, texts were then manually edited for any deficiency resulting from the conversion process (e.g., it was very common to lose words or even lines when converting PDF files that include Arabic text into plain text format).

4.5.3 Corpus Size

There are debates over the minimum total number of words (tokens) necessary for a specialized corpus to be considered representative (Biber et al 1998 and Bowker and Pearson 2002), as the size depends largely on language, genre, research question and distribution of linguistic features that are being studied (Leech 1991, McEnery and Wilson 2001, Meyer 2002 and Sinclair 2004b). A corpus used to study general language is typically much larger than one used to study specialized language, as the latter usually aims to represent a particular section of language. There is a growing recognition of the value of smaller specialized corpora (Ghadessy et al 2001; O’Keeffe 2007). Difficulties in securing data of limited availability or inaccessible formats require researchers to settle for relatively small corpora (Hunston 2002; How and Kan

31 It is recommended to enable Arabic language support in the operating system used (Windows8 in my case) before processing an Arabic corpus so that WST will display the Arabic scripts correctly.
Furthermore, compiling larger corpora may be possible only for wider and well-funded projects, which is not the case in the current study. Ahmad and Rogers (2005), therefore, suggest that a corpus of 100,000 words would be a good starting point for corpus-based terminology, especially in a highly specialized discipline such as computers and the Internet.

In this study the focus was on selecting texts that cover the domain of information technology. Taking into consideration the fact that specialized corpora can be relatively small in size as well as the severe lack of relevant electronic Arabic resources available on the WWW (discussed above), the content of the corpora was controlled by the availability of reliable materials in MSA that represent the domain in question. This concern is especially problematic when creating the corpus for Morocco, Algeria and Tunisia, where materials related to IT are almost always in French rather than in Arabic. This also affected the size of the corpora in that I found adequate materials for one corpus but few for the other.

The need to obtain permission from copyright holders is another major factor that affected the size and the structure of the corpora. It was not possible to obtain permission to use the texts from all copyright holders. In many cases there was no response from the owner, although requests were sent several times. Therefore, I assumed — like many others did, e.g., Spoor 1999 and Cooper 2003 — that a vast majority of the documents published on the Internet are not protected by copyright and that many authors of these documents are happy to be able to reach as many people as possible (Indurlihya and Damerau 155). Moreover, Hilton (2001) states that “If the use of a work furthers progress in the sciences and the arts (i.e., if it promotes learning, knowledge, and the public good) and if its use will do relatively little harm to the
author's property rights, then it is not necessary to get the author's permission to use the work.” (50)

Although the size of the two corpora is limited because of the constraints discussed above, they are sufficiently large for the purpose of this study, as the aim is to represent a specialized section of the Arabic language (computers and Internet.) Obviously, small, yet clean and comparable corpora are more suitable for the purpose of this investigation than larger collections of general language web pages or articles.

4.5.4 Processing the Corpora

After collecting and compiling the corpora, the next step was to apply a methodology to identify, process and analyze them. This section presents the steps that were taken to process and analyze the corpora after they were built. It also explains how WST special features were applied in this analysis in order to obtain more reliable results. The processing of the study corpora consists of two phases: the semi-automatic extraction of terms and terminological analysis of candidate terms. The two phases will be explained in the following sections.

4.5.4.1 Semi-automatic Extraction of Terms

For languages that use a non-Latin script such as Arabic, full automatic processing and analysis is usually limited due to reasons discussed previously in this chapter. Consequently, the method I followed in corpus processing was not entirely automatic. The analysis of the two corpora also could be done only in terms of term frequency and concordance. Once the corpora had been compiled, a stop list with function words for both corpora was created to avoid counting high frequency functional words (see 4.4.1 above for discussion about the WST stop-
list function). The stop list included 92 words extracted from the first 100 words in the initial frequency list. It should be noted that the stop list did not include all function words, but only the ones that have had the highest frequency of occurrences in each corpora. The words included in the stop list are listed in Appendix II. This list was used for the two corpora, but each corpus was processed separately. The following screenshot (Figure 4.7) shows the word list before the generation of the stop list. It is evident that function words such as "فِي" fee, "مِن" men, "ور" waw and "و" and have the highest frequency of occurrences in the word list (2.809, 2.309 and 1.538 respectively).

![Figure 4.7: Word list before the generation of function words stop list](image)

4.5.4.1.1 Term Inclusion

After the creation of the stop list, it was uploaded into WST in order to produce a list of content words in each corpus. Once function and unwanted words were excluded by means of the stop list and the most frequent content words in the two corpora were identified, specialized
single-word and multi-word terms for computers and the Internet were manually selected from the created word list. The *Wordlist* function in WST (see 4.4.1 above) was used to obtain the frequency of the words in each corpus. The selection of terms was made according to the following criteria:

(1) Common terms from information technology (i.e., terms that occur more frequently than others), which could be used by both specialized and non-specialized users;

(2) Terms coined by different means of term formation in Arabic, such as translation, derivation, and borrowing (for more details about these strategies, see Chapter Two); and

(3) Terms that potentially have many synonyms. For the sake of simplicity, I concentrated only on specific terms for which there is no established standard equivalent or an accepted lexical unit in MSA and that thus pose a source of difficulty for Arabic technical writers and translators (e.g., new computer and Internet terms that are too novel to be listed in Arabic glossaries and dictionaries).

**4.5.4.1.2 Term Exclusion (Lemmatization)**

At first glance at the term list, I noticed the problem of the lack of stemming (lemmatization\(^{32}\)), i.e., the lists include individual entries for both the base forms and for their inflectional forms. For example, the term شبكة *network* appeared both in singular as شبكة *network* and plural شبكة *networks* forms in the list. One reason for this problem is the highly derivational

---

\(^{32}\) Lemmatization (Stemming) in corpus linguistics processing is the process of grouping together the different inflected forms of a word so they can be analyzed as a single item.
and inflectional nature of Arabic language\textsuperscript{33} (see 4.3.1.3 above), coupled with the absence of any reliable lemmatization utility for the Arabic language.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.8.png}
\caption{Lemmatization of different inflectional words}
\end{figure}

WST was used to lemmatize inflectional words. For example, in Figure 4.8, the terms تكنولوجيا the technology with the definite article ال the (freq. 166) and تكنولوجيا technology without the definite article (freq. 89) were lemmatized together (no. 22), which thus accounted for 255 occurrences. The term تكنولوجيا technology retained its original position, but it had been reduced to zero occurrence in the lemmatization procedure\textsuperscript{34}. The same process was performed on the term شبكة network and الشبكة the network. Several terms in the same lemma were left in their

\textsuperscript{33} Many Arabic researchers, such as Abdulkarim (1990) suggest dividing the term up and placing a space between the base of the term and its possible prefixes and suffixes. Thus, instead of treating شبكة network and شبكات networks as two occurrences, the researcher can separate the suffixات (the plural form in Arabic) from شبكة (the base form of the term) so that the computer software counts the base and its inflectional forms as one frequency of occurrences. However, this method is time consuming, especially when it comes to processing a hugely complex language from a morpho-syntactic viewpoint, such as Arabic where so many combinations of different inflections are possible.

\textsuperscript{34} For a detailed instruction on how to join together related entries by means of WST lemmatization function see Wordsmith Tools Step by Step (Scott 2010) http://www.lexically.org/wordsmith/step_by_step_guide_English.pdf.
separate entries. For example, the verb *yatasafah* to *browse* and the noun *motasafih* browser are left in separate entries, even though they are derived from the same base, because they are different parts of speech.

The corpora files were then used to manually identify and compare synonyms to determine whether there is any term variation, regardless of whether the variation occurs within the boundaries of one corpus or in separate corpora. Terms are considered synonyms if they have the same meaning and are assumed to represent the same concept. For example, the terms (برمجة متعددة المنصات، عبر المنصات، البرمجة المتعددة المنصات) are synonyms and equivalents to the English term *cross platform*. The focus was primarily on orthography (e.g., spelling variations, vocalization, etc.), morphology (e.g., variations related to inflectional phenomena such as number and gender and derivational variations), lexicon (e.g., synonyms used interchangeably), phonological variations, and acronyms and abbreviations. It should be noted that these types of variation do not change the meaning, as they refer to the same concept. After the terms were identified manually, term lists were created and used as a starting point for setting up a term categorization framework. The following section presents the terminological analysis of the term lists extracted from the two corpora.

### 4.5.4.2 Terminological Analysis

This phase started with gaining an overall view of the most frequent terms in C1 and C2 by creating key term lists, which included terms that had higher frequencies of occurrence in the original list. WST defines the lists on the basis of single lexical units by recognizing characters between spaces. In many circumstances, however, this identification is not accurate, as many technical terms are made up of more than one word. For example, the term *جهاز غير متزامن* jihaz
ghayr mutazamin *asynchronous device* is considered one terminological unit consisting of two words in English *asynchronous* and *device*. Separating the words leads to a change in term meaning. In Arabic, the problem is more complicated, because this term contains three words, i.e., جهاز ghayr not and متزامن mutazamin *synchronous*. WST treats these three words as three separate units, each has its own frequency occurrences.

To solve this problem, WST Concordance was used. WST concordance made it possible to browse the corpora and visualize the searched terms in their original contexts (see 4.4.2 above). This tool enabled me to determine whether the term in the lists was used as a single word unit or a multi-word unit by looking at its most frequent neighboring words. The following screenshot is an illustration.

![Figure 4.9: Arabic concordance using WST](http://www.lexically.org/wordsmith/step_by_step_guide_English.pdf)
The above screenshot shows the concordance of the occurrences of the term **information** in C1. It lists all the occurrences of the term **العلومات** (highlighted in blue by the software) with their collocations (highlighted in red), such as **ابنك العلامات** **information banks**, **استرجاع العلامات** **information retrieval**, and **أنظمة العلامات** **information systems**. However, this output is not in the right order due to the RTL nature of Arabic script (see 4.3.1.1 for details about this feature of Arabic script). The occurrences of **العلامات** **information** are displayed at the beginning of the line, whereas their collocations, such as the word **ابنوك** **banks** in line 3 and **استرجاع** **retrieval** in line 8, are displayed at the end. It seems that WST concordance does not take the bi-directional nature of Arabic into account when displaying results. Admittedly, when I double-clicked on one of the concordance lines in order to see the source text, the ordering issue disappeared and words were shown in the right order for some unknown reason, but it seems that WST retrieved and showed the texts in its original txt files, which were already saved as Unicode (see 452 above).

Once all term occurrences in C1 and C2 were identified, the absolute frequency of each term was then calculated for each corpus. The two corpora were treated separately so as to generate a distinct term list for each one. In order to avoid analyzing a very large number of terms, only terms that were directly related to the domain of computers and the Internet and that had a frequency of 10 or higher were selected as key terms for analysis. As an example, Table 4.10 shows the top ten terms on the word frequency lists in C1 and C2. Full listings of the most common terms that occurred more than 10 times can be found in Appendix III for C1 and Appendix IV for C2. C1 yielded a key term list of 63 terms, whereas the list for C2 was 77 terms. Only terms recorded in these two final lists were taken into account in the final analysis.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>الانترنت</td>
<td>antrnet</td>
<td>Internet</td>
<td>220</td>
<td>الانترنت</td>
<td>enternet</td>
<td>Internet</td>
<td>239</td>
</tr>
<tr>
<td>تكنولوجيا</td>
<td>teknulujya</td>
<td>technology</td>
<td>231</td>
<td>بروتوكول</td>
<td>brutukul</td>
<td>protocol</td>
<td>168</td>
</tr>
<tr>
<td>حاسوب</td>
<td>hasub</td>
<td>computer</td>
<td>259</td>
<td>كمبيوتر</td>
<td>kumbutr</td>
<td>computer</td>
<td>151</td>
</tr>
<tr>
<td>شبكة</td>
<td>sabaka</td>
<td>network</td>
<td>235</td>
<td>شبكة الإنترنت</td>
<td>sabakat alentrnt</td>
<td>Internet network</td>
<td>139</td>
</tr>
<tr>
<td>HTTP</td>
<td>HTTP</td>
<td>HTTP</td>
<td>193</td>
<td>شريط الأدوات</td>
<td>saret aladwat</td>
<td>tool bar</td>
<td>122</td>
</tr>
<tr>
<td>كمبيوتر</td>
<td>kumbutr</td>
<td>computer</td>
<td>185</td>
<td>حاسوب</td>
<td>hasub</td>
<td>computer</td>
<td>99</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>antrnaat</td>
<td>Internet</td>
<td>118</td>
<td>الويب</td>
<td>web</td>
<td>web</td>
<td>97</td>
</tr>
<tr>
<td>بطاقة الذاكرة</td>
<td>betaqat althakera</td>
<td>memory card</td>
<td>97</td>
<td>مربع حوار</td>
<td>muraba’ hewar</td>
<td>dialogue box</td>
<td>83</td>
</tr>
<tr>
<td>Windows</td>
<td>Windows</td>
<td>Windows</td>
<td>81</td>
<td>لوحة المفاتيح</td>
<td>lawhat almafateh</td>
<td>key board</td>
<td>72</td>
</tr>
<tr>
<td>HTML</td>
<td>HTML</td>
<td>HTML</td>
<td>68</td>
<td>القرص الصلب</td>
<td>algurs alsalb</td>
<td>hard disk</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 4.10: The highest frequent terms in C1 and C2

As shown in Table 4.10, there are some similarities and differences between the C1 and C2 frequency lists. C2 has more terms than C1. The term **الانترنت antrnet** Internet, which came at the top of the frequency list in C1 with 220 occurrences, is also the most frequently used in C2 (239 occurrences) but with hamza added below the alif\(^{36}\), followed by تكنولوجيا teknulujya technology in C1 (231 occurrences) and بروتوكول brutukul protocol in C2 (168 occurrences). The total frequency of حاسوب hasub computer was 259 in C1 but only 99 in C2, while its loan form كمبيوتر kumbutr computer occurred 185 times in C1 and 151 in C2. This initial analysis produced the expected finding that Arabic writers tend to use loan terms even when a suitable Arabic term already existed. Further discussion is presented in Chapter Five. Abbreviated foreign terms such

\(^{36}\) The difference between these two forms of the term **الانترنت Internet** is discussed in detail in the next chapter (Chapter Five).
as HTTP and HTML are among the most frequent terms in C1. The former has 193 occurrences, whereas the latter has 68 occurrences. This is because of the abovementioned fact that computers and Internet discourse in countries representing C1 (i.e., Morocco, Tunisia and Algeria) are almost always represented in French (see Chapter Three).

After collecting all the key terms (terms that have high frequency of occurrence of 10 or more), concordance analysis was carried out. Terms in the C1 list were inspected and compared with those identified in C2. Cross-terminological aligned tables were compiled by consistently linking terms in one corpus with their synonyms elsewhere. The objective here was to visualize terminological relationships between terms from different regions so as to carefully examine, describe and classify the specific objects of this comparison. Comprehensive reference tables were then worked out for variations or similarities that were found, then the terms were listed in these reference tables. Table 4.11 is an example:

<table>
<thead>
<tr>
<th>Category</th>
<th>Arabic term</th>
<th>Transliteration</th>
<th>Meaning</th>
<th>C1 Freq.</th>
<th>C2 Freq.</th>
<th>texts</th>
<th>C2 Freq.</th>
<th>texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>LV</td>
<td>العتاد</td>
<td>al’atad</td>
<td>hardware</td>
<td>27</td>
<td>8</td>
<td>13</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>MV</td>
<td>حاسب آلی</td>
<td>hasib aly</td>
<td>computer</td>
<td>15</td>
<td>6</td>
<td>79</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>ORV</td>
<td>انترنتت</td>
<td>antrnyt</td>
<td>Internet</td>
<td>318</td>
<td>10</td>
<td>nil</td>
<td>nil</td>
<td></td>
</tr>
<tr>
<td>AV</td>
<td>شبكة خاصة</td>
<td>Sabaka khasa</td>
<td>VPN</td>
<td>45</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11: Comprehensive reference of terms and their variation categories

As shown in the above example (Table 4.11), the tables classified terms according to the source region from which they were taken (either C1 or C2), the number of occurrences (frequency) and
the number of texts in which the term occurred. The tables also included a classification of terms according to the type of terminological variation—for example, terminological variation from linguistic perspectives such as orthographic variation (ORV), lexical variation (LV), morphological variation (MV) and abbreviation (AV); terminological variation from a colonial perspective (i.e., terms based on choices made in accordance with the dominant foreign language, either French or English); and variation from a geopolitical perspective (i.e., terms based on relevant communication problems and the current concern of Arab countries to build up their individual political identity).

In order to facilitate analysis, a comparative analysis for the two corpora was developed for comparing the relative frequency of occurrence of computer and Internet terms in order to measure how different their treatment was on the basis of the sort of terms used in the two corpora. The lists created from C1 and C2 were thoroughly examined, and the extracted terms were presented in accordance with their frequency in the analyzed corpora. The results from this analysis were compared manually without fully depending on WST. In different cases, however, I had to return to the original lists in WST to check and confirm the analysis of certain terms. The advantage of conducting this comparison is that it not only helps evaluate terms against any one of their synonyms, but it is also useful for visualizing the evolution of certain terms. The results of the analysis of these two compiled corpora will be extensively reported in Chapter Five.
Two factors merit noting before concluding this chapter:

(1) Although frequency counts were worked out and reported in the study of terminological variation detected in the data, only descriptive statistical analysis was attempted throughout the discussion. Since the corpora are relatively small in size, it can be assumed that a statistical analysis of this small data set would be neither truly informative nor valid.

(2) Since the study is not concerned with terminology evaluation (rather, it is an empirical descriptive study), no attempt has been made at a comprehensive evaluation of terminology, nor does it stipulate which term should be used (i.e., there was no discussion of which term to use out of a number of synonyms). Instead, I analyzed the origins and formation of these terms and how they might have evolved.

Finally, several explanations and interpretations of these terminological variations were proposed in the light of the different foreign language orientations that are believed to have triggered them. A paradigm of terminological variations and their causes was built up. These causes potentially include the discursive nature of computer and Internet terms, some sociolinguistic factors related to the present status of diglossia in Arabic, the influence of foreign language hegemony, the exceptional richness of Arabic synonyms (which may lead to different Arabic equivalents for the same term), and political factors such as the absence of official transnational institutions that have the power to standardize thousands and thousands of new concepts in different fields of knowledge, particularly in the information technology arena. More general conclusions and recommendations were later drawn from this paradigm.
4.6 Summary and Conclusion

This chapter gives a general overview of the methodology adopted in creating two monolingual Arabic corpora of the special field of computers and the Internet with the aim of studying terminological variation in Arabic.

It can be concluded that corpus analysis tools are not user-friendly when it comes to languages with non-Latin script such as Arabic and Hebrew. There have been many problems in achieving ease and reliability in carrying out simple processing with these languages. To the best of my knowledge, there is still no tool available for reliable processing of Arabic script, mainly due to the unique script and complex structure of the Arabic language itself. It is very important, therefore, to extend the functionality of these tools to include the particular features of Arabic script such as BiDi, complex morphology, and vocalization in order to provide better support to users. Enriching words with some linguistic knowledge such as (POS\(^{37}\)) tags may resolve most morphological and structural problems, and simple recognition of Arabic conventions by software designers at an early stage of the development process can prevent many issues such as BiDi from becoming overwhelming challenges at subsequent phases. Unfortunately, one of the many difficulties facing Arabic corpus developers is the lack of adequate tools for encoding the markup in order to simplify Arabic word recognition and easily identify contextual meaning.

The methodology proposed here does have limitations, such as the availability of representative specialized Arabic texts that deal with computers and the Internet, which may affect the generalizability of the findings and conclusions. Nevertheless, it is hoped that the study will provide Arabic translators with a basis for determining the origins and formation of

\(^{37}\) e.g., the Qur\’anic Arabic Corpus developed by the Arabic Language Computing Research Group at Leeds University, see ‘section 4.3 Arabic corpora’ above.
technical terms in general and how they might have evolved, which may pave the way towards better theoretical and practical terminology work both horizontally and vertically among different systems used in different organizations and different regions. These consistent terminological strategies can help both individuals and companies, as they have strong motivation to achieve efficient communication and to cope with the current terminological explosion, especially in the arena of technical discourse, where standard Arabic terminologies are lacking.
Chapter Five

Patterns of Regional Variation in Arabic Terminology
5.1 Introduction

One of the many phenomena encountered when representing Arabic terminology is terminological variation, as concepts are often lexicalized by means of different terms. This is a phenomenon that has been the target of many corpus-based studies in other languages, such as Spanish (Freixa 2001), Japanese (Yoshikane et al. 1999) and Catalan (Freixa 2001), but English and French are the two languages that have been most widely analyzed (Daille et al. 1996 and Jacquemin and Royaute 1994). This chapter aims at giving a good picture of the phenomenon in Arabic terminology through the comparison and analysis of the two corpora on computers and the Internet generated from the two groups of Arabic-speaking countries as described in Chapter Four. The main concern of this chapter is to confirm or refute the hypothesis underlying this study, which assumes that the complexity of the multilingual situation in the ANAC is one of the main determining factors behind term variation and term formation. Although there are many different viewpoints that attempt to explain the causes of Arabic terminological variation, this current multilingual situation and the hegemony of foreign languages (see Chapter Two) are expected to show the strongest influence on technical term formation and use in the AW in general and the ANAC region in particular.

The chapter is organized as follows. Section one discusses several orthographic variations that were detected in the analyzed corpora using WST. It also examines non-linguistic factors behind these variations, such as the influence of foreign languages, as well as linguistic factors, such as the omission of short vowels and the application of the hamza ( ﹪) Arabic grammatical

38 In this study, three main factors which lead to Arabic terminology variation have been proposed: the discursive nature of the computer and Internet domain; the present status of diglossia in Arabic, and most importantly the influence of foreign language hegemony. More factors will be presented in Chapter Six, including political, economic and theoretical factors.
rules. Section Two presents terminological variations at the morphological level. It also discusses the influence of French and English borrowing on the terminology of computers and the Internet. The issues of grammatical gender and number, as well as phonological issues with reference to morphology, are also included. The next two sections focus on terminological variations in abbreviations and neologisms. These different kinds of variation are reviewed with examples collected from the study corpora and illustrating the phenomena in question.

The chapter concludes that it has been possible to identify substantial term variations that have resulted in consolidating the influence of foreign languages, especially French and English, as the languages of communications and work in the area of computers and the Internet. Thus the hypothesis presented in the beginning of this study and recalled in this introduction is supported.

To a certain extent, the data also support the hypothesis that terminology vitiation in computers and the Internet is the result of the existing dialectical variation among countries of the ANAC region (see Chapter Two). Translators and technical writers are apt to adopt local dialect roots for coining new terms if no other alternatives are found in either MSA or foreign loaned terms.

5.2 Results and Discussion

Term variation is a natural phenomenon which, far from being avoided, is widespread in specialized discourse (Freixa 2006, cited in Fernandez et al. 49) and which occurs as part of the terminology formation process. In theory, terms should be mono-referential (Spasic et al. 245), i.e., there should be a correspondence between concepts, with one term for one concept. The observation of the corpora in this study, however, shows that most Arabic terms are homographic, which means that the same concept corresponds to many terms. Terminological
variations that appear in the study corpora can be roughly classified into orthographical variations, morphological variations, lexical variations, abbreviations and neologisms. Each variation is caused by different factors, to name a few: dialectical, inter-lingual, functional, cognitive and discursive factors (Freixa 55). Each of these factors leads to a particular problem for term recognition. However, it is dialectical (different Arabic dialects) and inter-lingual (influence of foreign languages) factors that have contributed the most significantly to the phenomenon of Arabic terminology variation.

5.2.1 Orthographic Variation

Orthographic variation includes the use of Arabic vocalization (e.g., the presence or absence of diacritics or hamza (ء)), punctuation (e.g., the presence of a hyphen), as well as number (e.g., satellite network and satellites networks). These variations produce terminological units or synonyms that correspond to the exact same concept. Figure 5.1 illustrates orthographic variation where the same lexical sense (concept) is represented by different written representations (lexical entry):

![Diagram of orthographic variation](image)

Figure 5.1: Orthographic variation
For instance, the word ‘Internet’ is usually written as انترنت antrnt (without hamza ء) in the Middle East, in contrast to North Africa, where it is written انترنت antrnt (with hamza ء above the (!) alif. There is a reason behind this: in the Middle East the word “internet” is a transliteration from English, whereas in North Africa the transliteration of the French word for Internet is used. However, this orthographic variation involving the writing or omission of (ء) hamza above or below (!،أ) alif, as well as the use of bare alif (!) without hamza, is quite common in Arabic writing and, in many cases, goes unnoticed by most Arabic readers (Buckwalter 32). Therefore, this usage might not be regarded as orthographic variation by many Arabic scholars (e.g., Buckwalter 33) and can be viewed as non-problematic or treated as a typographical error. For that reason, the matter of orthographic variation attracts very little attention from Arabic researchers.

Nevertheless, these types of errors or mere omissions definitely affect the retrieval accuracy of information retrieval systems, as inflected words can have a significant impact on the retrieval effectiveness of any information retrieval system (Nenadic et al, 3). Therefore, submitting انترنت Internet with bare alif as a search query results in finding a different number of occurrences from those for انترنت with hamza above or below the alif انترنت entrnt. Consider the following table, which indicates different frequencies for the term انترنت Internet with and without hamza (ء):

<table>
<thead>
<tr>
<th>English term</th>
<th>Arabic term</th>
<th>Transliteration</th>
<th>Type of variation</th>
<th>C1 Freq.</th>
<th>C2 Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet</td>
<td>انترنت</td>
<td>anternet</td>
<td>no hamza</td>
<td>241</td>
<td>105</td>
</tr>
<tr>
<td>Internet</td>
<td>انترنت</td>
<td>anternet</td>
<td>hamza above alif</td>
<td>74</td>
<td>19</td>
</tr>
<tr>
<td>Internet</td>
<td>انترنت</td>
<td>enternent</td>
<td>hamza below alif</td>
<td>94</td>
<td>239</td>
</tr>
</tbody>
</table>

Table 5.1: Different frequencies for different spellings
Table 5.1 above shows different results for each of the three word forms, namely Internet with bare alif، انترنت with hamza above and انترنت with hamza below the alif. The reason for these variant retrievals is the absence or presence of hamza on or below the alif. The most frequent instance is انترنت with bare alif (241 for occurrences in C1 and 105 in C2), which means that this spelling of the word is more popular than the other two. This is because most Arabic texts on the Internet are introduced without diacritics, including hamza.

5.2.1.1 Orthographic Variation and Normalization

To handle this problem in Arabic language processing systems and improve information retrieval, researchers (e.g., Nenadic et al, 2004) implemented a normalization strategy (see Chapter Four for a detailed discussion). The initial (۱) alif in the words انترنت and انترنت with a hamza above or below (۱ and ۲) is replaced with simply an alif (۱) with no hamza. They also normalize the alif madda (۳) with a ‘bare alif’ (۱).

The analysis of the two corpora indicates that there is a tendency among Arabic writers in general to normalize words that include ء hamza. Table 5.1 shows that there is a clear difference between the frequency of occurrences of انترنت anternet with hamza above the initial alif and انترنت entrnet with hamza below the alif. While the first appeared 74 times in C1 and only 19 times in C2, the second appeared 94 times in C1 and 239 in C2, which has the highest frequency. An explanation for the higher frequency of انترنت entrnet as compared with انترنت anternet is that the first might be considered the appropriate loan as far as the pronunciation of the term Internet in the original foreign language is concerned, which points to a prevalence of English or French as the influencing language.
Nenadic et al. have demonstrated that normalization does result in significant improvements in the retrieval effectiveness of Arabic information retrieval systems. Table 5.2 illustrates some additional examples extracted from my corpora:

<table>
<thead>
<tr>
<th>English term</th>
<th>Arabic term</th>
<th>Normalized term</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>internet</td>
<td>أنترنت/إنترنت</td>
<td>انترنت</td>
<td>antrnt</td>
</tr>
<tr>
<td>spam</td>
<td>الرسائل التلفظية</td>
<td>الرسائل التلفظية</td>
<td>alrasayl altaiafulya</td>
</tr>
<tr>
<td>mouse</td>
<td>فأراة</td>
<td>فأراة</td>
<td>fara</td>
</tr>
<tr>
<td>mother board</td>
<td>اللوحة الأم</td>
<td>اللوحة الأم</td>
<td>allwaha alum</td>
</tr>
<tr>
<td>Ethernet</td>
<td>إيثرنت</td>
<td>إيثرنت</td>
<td>ethrnt</td>
</tr>
</tbody>
</table>

Table 5.2: Term normalization examples

These variations may also occur due to the missing short vowels, which are represented orthographically using the diacritics (فتحة fatha, ضمة dhama, كسرة kassra, etc.). Typically, these diacritics are rarely used in MSA electronic corpora and almost never in descriptive Arabic dialect corpora (see Chapter Two for a detailed discussion about the Arabic vocalization system). Al-Sabbagh and Girju (283) have recently supported the position that spelling variations due to the lack of phonological diacritics have always been claimed to be a problem in Arabic corpora. In fact, this is one of the many challenges in processing Arabic electronic corpora (for further discussion about phenomena associated with spelling variation in Arabic, see in particular Habash et al 2011 and Dasigi and Diab 2006). The first analysis of this study indicates that the absence of diacritics (discussed in Chapter Four) made processing and analyzing the study
corpora far from easy, due to the high degree of ambiguity in the Arabic texts. The analysis of the corpora reveals many words that have various valid analyses. For example, the word حاسب hasab below without diacritics is highly ambiguous, because it might have at least three valid analyses:

1- حاسب hassib  computer

2- حاسب hassab  request account

3- حاسب hassib  counter (person who counts)

WST treated these words as one word without differentiating among the variations. Therefore, I had to perform a concordance search for each word in order to determine its exact meaning. Features of Arabic such as lack of capitalization and complex word structure might also contribute to such ambiguity, but it is the absence of diacritical short vowels that causes the most problems.

5.2.1.2 Orthographic Variation and Foreign-language Influence

As noted initially in this study, terminology in Morocco, Algeria and Tunisia is strongly affected by the French language, which has a significant grip on terminology choices in spite of the long process of Arabization. These countries have been influenced by French for more than a century since the initial French colonization of Algeria in 1827. Hence Arabic in this part of the AW has taken over a certain number of phonological features from French, which has also had its effect on spelling. In Egypt and Libya, on the other hand, the rise of English terms in technical discourse is also noticeable, since English is commonly perceived as the language of science and technology.
Due to the influence of the French language on speakers from Morocco, Algeria and Tunisia, writers often adapt French phonological features by the insertion of certain sounds; thus the vowel (ي) ya is added to the loan term 'انترنت' Internet. In the corpus collected from Morocco, Algeria and Tunisia (C1), the placement of the long vowel (ي) is a noticeable phenomenon. Analysis reveals that the addition of the long vowel (ي) corresponds to the French pronunciation of the term. Writers from this region prefer to infix (ي) in the term انترنت, while Egyptians and Libyans do not, because this shift does not reflect English pronunciation. The frequency of the term انترنت antrnt without (ي) in C1 is (nil), and in C2 it is 123, whereas the frequency of the term انترنت with (ي) between the last two characters is 318 in C1 and (nil) in C2. The frequency of انترنت with an extra أ alif between and the final ت is 134 in C1 and (nil) in C2. Furthermore, the term مودم مودم has been adopted with the (ي) vowel inserted in the root to correspond in sound to French. It is spelled in C1 as مودم مودم (frequency 15) and in C2 as مودم مودم (frequency 30). This difference between the two terms (인터넷 internet and مودم modem) can be explained by proposing that each represents a different foreign-language reference. The long vowel (ي) in terms such as انترنت انترنت and and the long vowel (ا) in انترناات are infixed as the result of conscious or unconscious assimilation and emulation of the French pronunciation of the two terms, which is different from how the same terms are pronounced in English. Similarly, the phoneme قوقل (ق) gugle Google, which occurred 2 times in C1 and 30 times in C2, is replaced with غوغول (غ) Google ghoghl in C1 (12 occurrences in C1 and 0 occurrence in C2) in accordance with speech features of C1 and C2 speaking communities (see the discussion in Chapter Two about changes in the phoneme (ق) qaaf among Arabic regions).
This analysis supports the thesis that there are two different language sources for terminology: a French source in Morocco, Algeria, Tunisia, and an English source in Libya, Egypt and almost all of the rest of the AW. The pace and scope of terminology formation has depended largely on which of these languages is more dominant and has more power than the other.

5.2.1.3 Orthographic Variation and Brand Names

Another aspect of inconsistency is the diverse orthographic rendition of some technical brand names, which resulted from the fact that standardized transliteration systems still do not exist in Arabic. Transliteration is the process of rendering the letters of one alphabet in the letters of another with a different alphabetical system (Pinchuck 1977). In the age of globalization,

---

39 Transliteration or transcription of Arabic scripts into Latin script is a complicated problem even in Arabic general language discourses. One typical example is the word Qur’an, which has one spelling in Arabic القرآن, but tens of spellings in other languages that use Latin script. To name a few: Qur’an, Koran, Quran and Kuran.
many international brand name products and inventions have found their way into the Arabic language. Most of them have in fact been transliterated from the two main sources from which Arabic draws its technical and scientific terminology, i.e., English and French. There are no Arabic standardized transliterated forms for many of these brand names. Al-Qinal explains that:

[T]hose who undertook the task of translating foreign books into Arabic or those who came into contact with speakers of other languages had no preset rules for Arabicizing foreign words. The transliterated form of a given loanword was, thus, in concordance with their best knowledge of its pronunciation. At times, they were not adequately fluent in the source language, and, therefore, the transliteration form of a given loanword may be the end product of a mispronunciation rather than any real phonological or morphological modifications (3).

Other Arabic scholars—for example, Al-Karuri (167)—oppose this method and call for the use of other means, such as derivation.

The analysis of the corpora reveals this problem in Arabic texts, which include many transliterated brand names whose spelling in general tends to be inconsistent. These differences were present in the two corpora, as illustrated in Table 5.4:
<table>
<thead>
<tr>
<th>Brand name</th>
<th>C1</th>
<th>Ferq.</th>
<th>Transliteration</th>
<th>C2</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPad</td>
<td>أیباد</td>
<td>25</td>
<td>ibad</td>
<td>5</td>
<td>a-bad</td>
</tr>
<tr>
<td>Excel</td>
<td>اکسل</td>
<td>21</td>
<td>aksl</td>
<td>12</td>
<td>aksyl</td>
</tr>
<tr>
<td>Facebook</td>
<td>فاس بوك</td>
<td>47</td>
<td>fasbok</td>
<td>19</td>
<td>faysbok</td>
</tr>
<tr>
<td>Explorer</td>
<td>اكسبلورر</td>
<td>22</td>
<td>aïksblurar (without Alif)</td>
<td>30</td>
<td>aksblurar</td>
</tr>
<tr>
<td>Windows</td>
<td>ونديواز</td>
<td>87</td>
<td>wandwaz</td>
<td>59</td>
<td>winduz</td>
</tr>
<tr>
<td>Microsoft</td>
<td>مايكروسوفت</td>
<td>21</td>
<td>mykrusuft</td>
<td>42</td>
<td>maakrusuft</td>
</tr>
</tbody>
</table>

Table 5.4: Variations in borrowed brand names

Foreign brand names and proper names in general, as many contemporary linguists (e.g., Al-Shalabi, et al 2009) stipulate, should be adopted according to their pronunciations in the source language, because most of these names are used worldwide irrespective of a given language. Table 5.4 shows that there is no great difference between the two corpora as far as the frequency of adopting foreign brand names is concerned. The most frequent name in C1 is ونديواز Windows with 87 occurrences, whereas its synonym ونديوز winduz occurred 59 times in C2. The brand name اكسبلورر Explorer occurred 22 times in C1 and اكسبلورار aksblurar (with extra alif) occurred 30 times in C2. At the lowest point, the brand name اکسل اكس لل Excel occurred 21 times in C1 and اکسل aksyl (with ي) occurred 12 times in C2. The biggest difference occurred with IPad and Microsoft, where أیباد ibad occurred in C1 20 times more often than a-bad in C2 and مايكروسوفت aïksblurar occurred 21 times more often than ميكروسوفت maakrusuft (without أ after the initial م) in C2.
In many instances, however, these brand names are written in both Arabic and the foreign language (English/French), while in others they are transliterated in Arabic alphabet alone. For example, in Figure 5.2 the term *Internet Explorer* is written in English as ‘Internet Explorer’ and as a borrowed term Интернет Explorer Internet Explorer, as well as its common equivalent in Arabic مستكشف الإنترنت mustakshif al antrnt Internet Explorer. In such cases the colonial language is acting as an agent of modernization and/or westernization, imposing a foreign concept that elicits a terminological response or multiple responses.

![Figure 5.2: Borrowed brand names written in both English and Arabic](image)

According to Ferguson (264), this makes for a kind of ambivalence or even split personality, individually and collectively. This might also indicate that writers are not entirely comfortable with these borrowed foreign terms and do not want to accept them as natural parts of the Arabic lexicon. The use of foreign terms next to the Arabic equivalent might be associated with the fact that people borrow words simply because these foreign words might be easier to use than established MSA equivalents such as مستكشف الإنترنت mustakshif al antrnt Internet Explorer above, considering the universal acceptance of foreign terms as well as people’s competence in MSA as discussed in Chapter Two.

There can also be cases where prestige plays a role in using the English or French term. Hence it has long been accepted that languages of lower socio-economic and political prestige borrow
from languages of higher socio-economic and political prestige. In a study of lexical borrowing between French and English in Canada, Grant-Russel states that the greater influence of one language over another can be accounted for by presuming a ‘prestige differential’ (482). Another factor, which may influence the acceptance of English and/or French terms, might be the acceptance of these terms by Arabic media, as it has a major involvement in Arabic term creation and dissemination.

5.2.2 Lexical Variation

The analysis of the corpora reveals that lexical variants are numerous, ranging from terms used more or less daily (such as *fax*—البريد المصور albard almusawar in Egypt, but الناشر النافع alnasir alnaf in Morocco—and *mother board* اللوحة الرئيسية al-lawha alraisya in Egypt, but اللوحة الأم al-lawha alum in Morocco) to more specialized but nevertheless familiar computer and internet concepts (such as *fiber optic*, which is الألياف البصرية al-alyaf albasaray in Egypt but الليفية البصرية la-lefiya albasaray in Morocco). However, the concept of lexical diversity is ill-defined, because in Arabic literature no fine distinction is made between lexical diversity and lexical richness, since “Arabic is famous for its exceptional lexical richness” (Beeston 102). Many Arabic scholars, such as Ebn Faris, refer to this lexical diversity and the existence of synonyms in Arabic as proof of Arabic lexical richness (see Versteegh 1990). However, lexical diversity is only one factor of the multi-dimensional feature of lexical richness. Read (62) proposes other factors such as lexical sophistication, number of errors, and lexical density.

Most often, lexical diversity can be measured through a type-token ratio (TTR), which refers to the total number of distinct words in relation to the total number of running words, i.e., comparing the number of different words (types) with the number of total words (tokens) in a corpus. Generally speaking, a lower TTR reflects greater consistency and a higher TTR, greater
lexical variety and less consistency. Statistics in Table 4.9 show that C1 has more lexical
diversity than C2, as C1 has a higher TTR than C2 (see the discussion of Table 4.9, Chapter
Four).

Synonym terms are those term variants that are considered semantically coincident but
formally different. The use of any of these synonym terms may vary depending largely on
geographical distance. The terms شبكة الإنترنت šabakat al-antrnt Internet web (which appears 38
times in C1 and 130 times in C2), شبكة الويب šabakat al-web Web (15 times in C1 and 5 times in
C2), شبكة الـشبكة العنكبوتية al-šabaka al-ankabotya spider web (38 times in C1 and 28 times in C2),
شبكة العالمية al-šabaka al-alamya world web (1 time in C1 and 30 times in C2) and شبكة الـشبكة العنكبوتية العالمیة
alšabaka alа’nkabutya ala’lamya world spider web (5 times in C1 and 28 times in C2) are all
considered synonyms for the concept Internet. Note that there are three translation equivalents
شبكة الإنترنت, شبكة العالمیة, شبكة الـشبكة العنكبوتیة العالمیة, شبكه الـشبكة العنكبوتیة
(šabaka al-ankabotya, šabaka al-alamya, šabaka al-ankabotya al-alamya), two transliterations of the foreign term web (شبكة
الويب, شبكة الويب, شبكة الـشبكة العنكبوتیة), as well as one loan (شبكة
الـشبكة العنكبوتیة الـانترنت, šabakat al-antrnt). Synonyms, however, are desirable only for familiarizing readers with
ordinary terms and words in general language, not for technical terms (de Waard and Nida 140).
Terminological problems arise as a result of this synonym diversity, because the Arabic reader
might wrongly assume different meanings for each alternative term.
<table>
<thead>
<tr>
<th>Arabic term</th>
<th>Transliteration</th>
<th>Meaning</th>
<th>C1 Freq.</th>
<th>C2 Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>الشبكة العالمية</td>
<td>alṣabaka Ala’lamya</td>
<td>World Web</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>الشبكة العنكبوتية</td>
<td>alṣabaka ala’nkabutya</td>
<td>Spider Web</td>
<td>38</td>
<td>28</td>
</tr>
<tr>
<td>شبكة الويب</td>
<td>šabakat alwib</td>
<td>Web net</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>شبكة الإنترنت</td>
<td>šabakat alintrnt</td>
<td>Internet web</td>
<td>58</td>
<td>130</td>
</tr>
<tr>
<td>شبكة الويب العالمية</td>
<td>šabakat alwb ala’lamya</td>
<td>World web</td>
<td>33</td>
<td>10</td>
</tr>
<tr>
<td>الشبكة العنكبوتية العالمية</td>
<td>alṣabaka ala’nkabutya ala’lamya</td>
<td>World spider web</td>
<td>15</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5.5: Arabic synonyms of the English term ‘Internet’

The statistics in Table 5.5 above show that there is no great difference between the usages of these synonyms. All of them are used regularly, although some terms were used more than others—for instance, الشبكة الإنترنت šabakat alintrnt Internet web, which has the highest occurrence in both corpora, occurred 38 times in C1 and 130 times in C2. The least used synonym was شبكة الويب šabakat alwib Web net, which occurred 15 times in C1 and only 4 times in C2. The very high frequency of occurrences of الشبكة الإنترنت šabakat alintrnt Internet web indicates that it might be easier to use than the others, such as the three-word synonyms شبكة الويب العالمية šabakat alwb ala’lamya World web and الشبكة العنكبوتية العالمية alṣabaka ala’nkabutya ala’lamya World spider web.

Younis states that
وفي اللغة العربية ليس هناك مجال لتعريب مصطلح الإنترنت، فالوسيلة الوحيدة لإدخال هذا المصطلح إلى اللغة العربية هي استخدام إمكانات المرونة في هذه اللغة العظيمة لإدخال هذا المصطلح إليها، دون حاجة لاستخدام مصطلحات محفزة أو إعلامية مثل (شبكة) أو (شبكة عالمية) أو (شبكة عنكبوتية)... الخ. فالإنترنت وصلة أو وسيلة تواصلية ليست شبكة، فالاختلاف كبير بين الاثنين... الخ وكتابة الاستخدام الخاطئ، ولو كان في شكل مجازي يؤدي إلى اختلاف المفاهيم والمصطلحات وبيما يتوافر به مناخ الاختلاف في إيجاد مصطلح موحد.

In Arabic, there is no means for the Arabization of the term ‘Internet’. The only way to introduce this term to Arabic is to use the flexibility features of the language without the need to use motivating or media terms such as (network) or (global network) or (world wide web) ... etc. The Internet is a connecting and communicative vehicle, not a network, and the difference between the two is quite large .... Many inappropriate usages of the term, even pragmatically, lead to differences between the concepts and their implications as well as to the existing variant and non-normalized terminological environment. (Translation by the author)

In many cases, the two corpora display synonym richness40. Several terms (e.g., homepage) appeared in both C1 and C2 and were used interchangeably with their synonyms (saßhat albedia start page with 23 occurrences in C1 and 28 occurrences in C2, الصفحة الرئيسية alsafha alraesia main page with 17 occurrences in C1 and 13 occurrences in C2, الصفحة الأساسية alsafha al-asasia basic page with 8 occurrences in C1 and 3 Occurrences in C2) and can

40 While the terms ‘terminological variation’ and ‘synonymy’ have been used without a clear qualification of their meaning or definition (Perazzo 119), the term ‘synonym richness’ is used here to overcome the negative connotation that have been associated to the term ‘terminological variation’ in terminological studies.
be considered to have the same meaning. On several occasions, these synonyms were used inconsistently even within the same text by the same author. The follow table presents another example:

<table>
<thead>
<tr>
<th>Spam</th>
<th>Transliteration</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>الرسائل التائفية</td>
<td>al-rasal al-taqha</td>
<td>unimportant messages</td>
</tr>
<tr>
<td>الرسائل الإلكترونية التائفية</td>
<td>al-rasal al-alaktrunya atafha</td>
<td>unimportant electronic messages</td>
</tr>
<tr>
<td>الرسائل المهمّلة</td>
<td>al-rasal al-muhmala</td>
<td>neglected messages</td>
</tr>
<tr>
<td>الرسائل التطفلية</td>
<td>al-rasal al-tatafolia</td>
<td>irruptive messages</td>
</tr>
<tr>
<td>الرسائل المزعجة</td>
<td>al-rasal al-moz’ha</td>
<td>irritating messages</td>
</tr>
<tr>
<td>الرسائل المتطفلة</td>
<td>al-rasal al-mota’tafla</td>
<td>intrusive messages</td>
</tr>
</tbody>
</table>

Table 5.6: Arabic synonyms of the English term ‘spam’

Table 5.6 shows that the English term *spam* has 5 Arabic equivalents, which are regularly used in the C1 and C2 alike. One of the main reasons for this discrepancy might be the tendency of Arabic writers to use these terms interchangeably depending on what kind of effect they want to produce on their readers (e.g., words such as المزعجة al-moz’ha *irritating* and المتطفلة al-mota’tafla *intrusive* are more aggressive and thus might produce a more negative effect on readers than neutral words such as التائفه atafha *unimportant* and المهمّلة al-muhmala *neglected*.)

This discrepancy might also be attributable to the current lack of coordination among different agents (organizations, governments, translators, publishers, suppliers, trainers, and so forth). This is another challenge that leads to the underperformance of terminology work. No doubt, coordination is crucial when it comes to language planning and terminology...
standardization, as their development depends on investments being made by other agents. The process of terminology creation and adoption requires close collaboration among terminologists, subject specialists, linguists, translators, and educators (UNESCO). Arabic terminology work, however, seems to suffer from an ‘information silo effect’ that has led to a huge gap of communication and coordination. The concept of “information silos” means that each community of practice (COP), government agency, scholarly discipline, etc., owns a separate, private set of information, often with significant amounts of content duplication. Every COP stores its own information and does not share it with others. One of the many reasons for this isolation is the variation in political stands within and among ANAC countries; there are different types of regimes and governing styles. As Emery puts it,

Much of the duplication in terminology creation in Arabic can be attributed to geographical and political factors — the sheer size of the Arab world (and the problems of communications therein) and the concern of Arab states to build up their own particular educational systems and develop their own political identities. (85)

According to Fishman (216), language-related research generally requires close collaboration and integration of disciplines that are still too rarely in serious contact with each other.

Another reason for such variation is the fact that some of the newly absorbed technical and scientific concepts are too novel to Arabic culture, and so translators, terminologists, technical writers, journalists, etc. have been forced to use more than one term to describe them. For instance, when the term ‘flash memory’ was introduced to Arabic culture it was translated as رقیّة ذاكرة السریعة raqīqat aḍakāra asarī’a fast memory card before it was simplified as بطاقّة ذاكرة السریعة.
bīqa't aḍākīra *memory card*, because the first term did not enjoy the popularity of the second despite being included in many dictionaries as a standard term. The analysis of the study corpora show that bīqa't aḍākīra *memory card* occurred 97 times in C1 and 55 times in C2, whereas raqīqat aḍākīra asārī'a *fast memory card* did not occur in either C1 or C2.

Finally, it is perhaps safe to assume that these lexical variations do not hinder comprehension among speakers and readers from different dialect areas because they are used in daily life. When it comes to more specialized terms in the domain of computer and Internet, however, where terms should be mono-referential (Spasic et al 2003), the variations will definitely hinder the communication process. Failures of communication might occur because of the highly specialized nature of the domain, which requires accuracy and consistency.

5.2.2.1 Lexical Variation and Foreign-Language Influence

The influence of French on Arabic was apparent during the first phase of collecting texts for C1. Websites from which texts were used for the components of C1 were characterized by the excessive use of French terms, even though these websites were published in Arabic. This was not the case with websites used for collecting C2 data. This point is also recognized by Zakaria Sahnnon (Arab Newspaper, 8 June 2008), who maintains that...

وَبِتَصِفُ جَلَّ الْمُوَاقِعِ الْمُغْرِبِيَّةِ، صِنَافِهِنَّ أنَّ غَالِبَتَهَا تَسْتَعِمِلُ اللُّغَةِ الفَرْنَسِيَّةَ، كَمَا أَنَّهَا بِنَفْصِهَا الإِبْدَاعِ
وَالجَدِّيْدُ حَيْثُ يَلْبَحُّ عَلَى هَذِهِ الْمُوَاقِعِ تَكُرُّرُ المَعْلُوْمَاتِ الْمُوجُودَةَ فِي الْمُوَاقِعِ الْأَخْرَى، بَسْبُوبِ الْإِعْمَادِ عَلَى
المَوَاقِعِ الأَجْنِيِّبِ (وُكَالَاتَ الْأَلْبَانِاءِ، مَرَاكِزِ الْبَحْرَاتِ...الخٌ)
…When surfing Maghreb sites, we will notice that the majority of them use French. Many of them lack creativity, and they update and repeat information found in other sites because of their dependency on foreign sites (e.g., news agencies, research centers, etc.). (Translation by the author)

Given this reliance on foreign sources, there is an ever-present influence of foreign terms and phrasing on Arabic writers. It is possible to explain the lexical variations between the two sides of the ANAC region by proposing that each represents a different foreign language reference. The influence of French on the lexical level through literal translation of certain terms is apparent. Table 5.7 illustrates some of these variations. In the first example, the term العتاد al’atad (which literally means materials and is used as an equivalent for the English term hardware) is a literal translation of the French word matériel, which is commonly used in French as an equivalent for the word hardware. The term العتاد al’atad occurred 27 times in C1, whereas its synonym اجهزة aẖhïza occurred 33 times in C2.
<table>
<thead>
<tr>
<th>English term</th>
<th>C1 Term</th>
<th>Freq.</th>
<th>Transliteration</th>
<th>C2 Term</th>
<th>Transliteration</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>hardware</td>
<td>27</td>
<td>al’atad</td>
<td>اجهزة</td>
<td>33</td>
<td>aţhîza</td>
<td></td>
</tr>
<tr>
<td>dialogue box</td>
<td>21</td>
<td>‘ulbat alhewr</td>
<td>مربع الحوار</td>
<td>83</td>
<td>muraba’ alhwar</td>
<td></td>
</tr>
<tr>
<td>video projector</td>
<td>8</td>
<td>akis alfidyu</td>
<td>مسلط</td>
<td>3</td>
<td>mislai/mîwar</td>
<td></td>
</tr>
<tr>
<td>mother board</td>
<td>37</td>
<td>allawha Alaum</td>
<td>اللوحة الرئيسية</td>
<td>34</td>
<td>allawha alraysya</td>
<td></td>
</tr>
<tr>
<td>scanner</td>
<td>18</td>
<td>mashiâ thwaïa</td>
<td>ماسح ضوئي</td>
<td>12</td>
<td>masîh dhwaï</td>
<td></td>
</tr>
<tr>
<td>software</td>
<td>34</td>
<td>mantiqya</td>
<td>برامج</td>
<td>52</td>
<td>baramḥ</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7: Lexical variations among Arabic countries

It is obvious that there must be some historical factors behind this influence in this part of the Arab World. During the colonial period (1830-1950) the French language was imposed by the colonizer in order to make Morocco, Algeria and Tunisia completely dependent on France in cultural, political and economic terms. French was the only language allowed in education, government and formal business activities (see Section 2.5 in Chapter Two for more discussion about the role and the status of French). Another reason is the geographical proximity to France and the fact that the in-place technology favors the adoption of French terms rather than English. This attitude may be explained by several factors (discussed in greater detail in Chapter Two), including the increased prominence of French as a medium for modern science and technology, the current heavy dependence of economic and educational activities in this part of the ANAC region on France, and the prestigious status assigned to French as the language of social superiority and professional success. Consequently, Arabic technical language in the ANAC region...
region has absorbed many terms from French in order to meet its technical communication needs. Table 5.8 illustrates some lexical items related to general technology that came from French.

<table>
<thead>
<tr>
<th>French term</th>
<th>Arabic term</th>
<th>Transliteration</th>
<th>Equivalence</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>icône</td>
<td>أيقونة</td>
<td>ïquna</td>
<td>رمز</td>
<td>Ramz</td>
</tr>
<tr>
<td>dossier</td>
<td>دوسي</td>
<td>mudal</td>
<td>مجلد</td>
<td>mojalad</td>
</tr>
<tr>
<td>dynamo</td>
<td>دينمو</td>
<td>dynamu</td>
<td>مولد</td>
<td>muled</td>
</tr>
</tbody>
</table>

Table 5.8: Examples of French terms integrated into Arabic

All the above-mentioned terms are fully integrated into the Arabic lexicon, although they have Arabic equivalents that are rarely used, with the exception of the term أيقونة ïquna, icône, which does not have an Arabic equivalent.

5.2.2.2 Lexical Variation and Dialectal Influences

As mentioned earlier in this study, Arabic dialects differ substantially from one region to another in terms of phonology, morphology, syntax, lexical choices and semantics. The degree of difference depends largely on the specific dialects of Arabic involved and how far the dialects are from each other geographically (for an exhaustive discussion of the difference between MSA and regional dialects, see Chapter Two). Furthermore, it is quite natural for written language (including MSA) to be influenced by spoken dialects, as “spoken dialect is more basic than written language” (Lyons 11). Previous research indicates that colloquial terms are often the only option that exist for certain discourses when no other alternatives are found—for instance, with
terms for agriculture and food production in MSA (see Wilmsen and Youssef 2009). I have also
demonstrated (Chapter Three) that colloquial terms derived from Italian are the only ones
available in Libyan automotive discourse.

It can be safely assumed that many terminological variations can be traced back to
phonetic and phonological differences between MSA and Arabic dialects on the one hand and
among different regional dialects on the other. The comparison of the two corpora reveals that
certain sounds that are absent in certain dialects are replaced by other sounds. We have noted
inconsistencies in characters that exhibit dialectal variations, such as (ق) qaaf, (ث) ṭa and (ج) ḥīm.
The MSA consonant (ق) qaaf, for instance, is realized as a glottal stop in Egypt and the eastern
region of Libya, whereas in the Maghreb countries it is /g/ (see Chapter Two for a detailed
discussion about the phonological differences among the Arab countries). Writers who adhere to
these different dialects, therefore, render these phonological differences in the orthography as
follows:

<table>
<thead>
<tr>
<th>English term</th>
<th>C1</th>
<th>Transliteration</th>
<th>C2</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>mega bit</td>
<td>ميغا بيت</td>
<td>maga byt</td>
<td>ميغا بيت</td>
<td>maha byt</td>
</tr>
<tr>
<td>Ethernet</td>
<td>إيثرنت</td>
<td>iṯrnt</td>
<td>إسربنت</td>
<td>isarnnt</td>
</tr>
<tr>
<td>network topology</td>
<td>تولوجيا الشبكة</td>
<td>tbluḥya aşabaka</td>
<td>طوبولوجيا الشبكة</td>
<td>ṭubluḥya aşabaka</td>
</tr>
<tr>
<td>pedagogy</td>
<td>بيداغوجي</td>
<td>bedaguhyy</td>
<td>بيداغوجي</td>
<td>bedaghoḥyy</td>
</tr>
<tr>
<td>gaga</td>
<td>قيقة</td>
<td>qaquia</td>
<td>جيجا</td>
<td>ḥaiha</td>
</tr>
</tbody>
</table>

Table 5.9: Phonological variations among Arabic countries

172
Abu Al-Yayyib Al-lugawi (351-962), who supports this assumption, claims that sound substitution is due to differences among Arabic dialects, especially in cases where the two letters in question belong to the same phonetic category. This appears most recently in the case of phonologically incompatible consecutive letters, such as the example of the term تبولوجيا الشبكة tbullhya ašabaka network topology in the above table, where the letter ت ‘t’ makes pronunciation difficult and was therefore replaced by the letter ط ubluhya ašabaka.

According to Dasigi and Diab (320), the lack of orthographic standards in Arabic dialects affects the spelling of all kinds of terms. Another explanation for this is that people translating new terms are not conscious users of MSA or native speakers of the language, simply because Arabic (or MSA) does not have native speakers (see Chapter Two for a detailed discussion for this phenomenon). This would contribute to an improper adoption of foreign terms, instead of creating a new term or using existing terms in MSA to describe newly introduced concepts.

5.2.3 Morphological Variation

Like other Semitic languages, Arabic has a rich and complex morphology (see Chapter Three for a detailed discussion about Arabic morphology). This results in a very sparse distribution of forms (Dasiqi and Diab 321). For example, many words can be derived from the root حس ب hasaba, which means compute in English. Table 5.10 illustrates morphological variations of the word حاسوب hasab computer detected in the two corpora:
The results in Table 5.10 show that 12 morphological variations were used in C1 and C2 with various frequencies. The term حاسوب has the highest frequency in the two corpora, with 259 occurrences in 10 texts in C1 and 99 occurrences in 10 texts in C2, followed by الحواسب which appeared 41 times in 7 texts C1 and 71 times in 7 texts in C2. While الحواسيب appeared 31 times in only 3 texts in C1, it appeared only 5 times in 2 texts in C2. Terms such as تHASAB to count with and احتسب to assume have the fewest occurrences in both C1 and C2, with 2 occurrences in 1 text for the former and only 1 occurrence in 1 text for the latter. Such morphological complexity poses a challenge to the development of accurate information retrieval output when working with automated information systems. The problem here is that each word input represents one or more potential tokens, each of which

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Transliteration</th>
<th>Meaning</th>
<th>C1 Freq.</th>
<th>Texts</th>
<th>C2 Freq.</th>
<th>Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>حاسوب</td>
<td>Hasub</td>
<td>computer</td>
<td>259</td>
<td>10</td>
<td>99</td>
<td>10</td>
</tr>
<tr>
<td>حاسب</td>
<td>Hasib</td>
<td>computer</td>
<td>26</td>
<td>5</td>
<td>47</td>
<td>6</td>
</tr>
<tr>
<td>حاسب الالي</td>
<td>hasib aly</td>
<td>computer</td>
<td>15</td>
<td>6</td>
<td>79</td>
<td>8</td>
</tr>
<tr>
<td>الحاسبات</td>
<td>hasibat</td>
<td>computers</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>الحاسب</td>
<td>hawasb</td>
<td>the computers</td>
<td>41</td>
<td>7</td>
<td>71</td>
<td>7</td>
</tr>
<tr>
<td>الحواسيب</td>
<td>hawasib</td>
<td>the computers</td>
<td>31</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>الحاسب</td>
<td>al-hasib</td>
<td>the computer</td>
<td>23</td>
<td>5</td>
<td>42</td>
<td>9</td>
</tr>
<tr>
<td>للحاسب</td>
<td>lil-hasib</td>
<td>for computer</td>
<td>24</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>بالحاسب</td>
<td>bil-hasib</td>
<td>by computer</td>
<td>19</td>
<td>4</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>كالحاسب</td>
<td>kal-hasib</td>
<td>like computer</td>
<td>26</td>
<td>3</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>تخاسب</td>
<td>tahasab</td>
<td>to count with</td>
<td>2</td>
<td>1</td>
<td>nil</td>
<td>nil</td>
</tr>
<tr>
<td>اخثساب</td>
<td>ihtasab</td>
<td>to assume</td>
<td>1</td>
<td>1</td>
<td>nil</td>
<td>nil</td>
</tr>
</tbody>
</table>

Table 5.10: Arabic morphological complexity
needs to be submitted individually for enquiry in order to obtain adequate results. For example, the input للحسابات لـ līl hasībat for computers can be considered as a single word token, yielding three morphological segments (līl ‘for’, hasīb ‘computer’ and aat ‘plural suffix’); the word كالكمبيوتر like the computer also yields three morphological segments (ka ‘like’, al ‘the’ and kumbutr ‘computer’); and so forth.

Furthermore, unlike English, the Arabic definite article (ال) al the is attached to words, creating compound forms that further complicate the manipulation of text in electronic corpora. For instance, the definite article (ال) al-hasib in الحاسب is not a separate word as in English the computer and many other Indo-European languages, but is appended to the word حاسب computer. In addition to (ال) the, conjunctions (such as و wa and والحساب واالحساب) (see Chapter Three), prepositions (للحساب līl hasīb for computer), and pronouns (حساسه hasībahu his computer, حاسبها hasībaha her computer, etc.) are attached to words, causing retrieval difficulties because a query that contains, for instance, the word للحساب līl hasīb for computer will have no match, or the output will differ from a query that contains any of the words listed in Table 5.10. Standard tools that have been developed for Indo-European languages, therefore, need to be modified to provide conflation capabilities that will account for predictable variation.

The term حاسب hasīb computer and its derivations are exclusively found in C2, whereas رتبة rattaba (literally meaning ordering things) is found only in C1, where French is the dominant foreign language. It is noteworthy that rendering the concept computer as رتبة rattaba returns to the original French term ordinateur, which – like its Arabic equivalent رتبة rattaba – carries the meaning of ordering things. This variation supports the assumption that each corpus represents a different foreign-language reference.
The direct loan كمبيوتر kumbutr for computer, however, is commonly found throughout the two corpora. This term has a frequency of 185 in 10 texts in C1 and 151 in 10 texts in C2. This high frequency of occurrences might be explained by the fact that this loan is informal and may appear in any kind of communication, either formal or informal, whereas its Arabic equivalent حاسوب hasub is more formal and restricted in standard communications. Furthermore, the universal frequency (i.e., the frequency of the term in the two corpora) (124) reflects the strong influence of foreign languages in general on terminology formation in the AW. It is worth noting that the English word computer has become an international word, as it occurs in several languages with the same or at least similar orthographic or phonemic form (see ISO 12620). This internationalism occurs as a result of simultaneous direct borrowing and acceptance from one language into another. Internationalisms frequently reflect Latin, Greek or English origins, but other languages, such as Arabic, French, Russian, Chinese and Japanese, have also contributed to the creation of internationalisms (ISO 12620).

Number and gender are also noticeable phenomena in terminology formation. The Arabic language is characterized by a strong tendency for agreement between elements. In matters of gender and number, these characteristics are expressed by a comprehensive system of affixes and inflections. In Arabic all nouns are classified into masculine or feminine. Feminine gender is often associated with a bound morpheme تاء التأنيث ta al-taneeth ‘feminine ta.’ Masculine gender, on the other hand, may be indicated with no suffix. For example, the word طابعة printer is a feminine noun, because ta marbota (ة) is attached to the end of the word. The dual41 is also formed by adding (ان) ‘ina’ to the end of masculine nouns and (بين) 41 Unlike English, which has just two numbers (singular and plural), Arabic nouns have three different kinds of number: singular, refers to one person or thing, dual refers to two and plural to more than two. While the singular is
'yan’ to the feminine nouns. The plural is formed by adding (ات) ‘at’ at the end of nouns. In all cases, the inflections can be irregular, as is the case in English and many other languages. The pattern of the singular noun is dramatically altered in certain cases. Table 5.11 illustrates the number and gender inflections of the word طابعة printer:

<table>
<thead>
<tr>
<th>Examples</th>
<th>Transliteration</th>
<th>Suffixes</th>
<th>English explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>طابعات</td>
<td>tabiāt</td>
<td>ات</td>
<td>printers/feminine/ plural</td>
</tr>
<tr>
<td>طابعتان</td>
<td>tabiātan</td>
<td>تان</td>
<td>two printers/feminine/dual</td>
</tr>
<tr>
<td>طابعة</td>
<td>tabiāh</td>
<td>ة</td>
<td>a printer/ feminine/singular</td>
</tr>
<tr>
<td>طابعين</td>
<td>tabeen</td>
<td>ان</td>
<td>typewriters/ masculine/ plural</td>
</tr>
<tr>
<td>طابعان</td>
<td>taban</td>
<td>ين</td>
<td>two typewriters/ masculine/ dual</td>
</tr>
</tbody>
</table>

Table 5.11: Gender and number inflection

The morpho-syntax of the reference language (French or English in our study) is a significant factor in determining term gender in Arabic. Whether a term is feminine or masculine tends to be affected by the gender of the equivalent foreign term. The use of feminine terms seems to be dependent on foreign language influence. For instance, the use of a feminine noun to refer to the word فأرة mouse is more frequent in C1 than its counterpart in C2. فأرة mouse (feminine) occurred 31 times in C1 and nil times in C2, while فأر mouse (masculine) occurred (12) times in C1 and 44 times in C2. Similarly, the feminine term مساحة ضوئية masiha thu’ya considered the base form of the noun, dual and plural have various morphological forms extended from the base form of the noun.
scanner occurred 18 times in C1 and nil times in C2, whereas the masculine term ماسح ضوئي scanner occurred only 4 times in C1 and 12 times in C2.

<table>
<thead>
<tr>
<th>Arabic term</th>
<th>Meaning</th>
<th>Frequency in C1</th>
<th>Frequency in C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>فأر</td>
<td>masculine mouse</td>
<td>12</td>
<td>44</td>
</tr>
<tr>
<td>فأرة</td>
<td>feminine mouse</td>
<td>31</td>
<td>nil</td>
</tr>
<tr>
<td>ماسح ضوئي</td>
<td>masculine scanner</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>ماسحة ضوئية</td>
<td>feminine scanner</td>
<td>18</td>
<td>nil</td>
</tr>
</tbody>
</table>

Table 5.12: Gender variation

This indicates that writers in Morocco, Algeria and Tunisia accept the use of the words فأرة mouse and ماسحة ضوئية scanner (feminine), which were coded in C1 based on the assimilation of the French words ‘la souris,’ and ‘échographe’ (feminine nouns in French), whereas the low frequency of the feminine words فأرة and ماسحة ضوئية in C2 indicates that they are not regularly used in Egypt and Libya. فأر mouse and ماسح ضوئي scanner (masculine) are more popular than the feminine terms فأرة and ماسحة ضوئية. Furthermore, in many cases this ء or ء‘ta marbuta’ that indicates feminine nouns is written as ء or ء ha without the two dots, which makes word retrieval even more difficult. This can be attributed to the fact that the ء or ء‘ta marbuta’ is rarely pronounced as ta ء or ء. It is often pronounced as ha ء or ء with or without the two dots.
The above comparison indicates that it might be interesting for further research to investigate the issue of gender sensitivity in specialized Arabic contexts, since the subject of gender in specialized term formation discourse has not been addressed in any great detail. Bowker (2001) acknowledges that “although there are numerous guidelines available for term formation, the issue of gender sensitivity is not typically addressed in these guidelines” (589).

Morphological variation also often appears in Arabic plural formation, as in, for example, Table 5.12, which shows that there are two plural forms to the word حاسب computer: حاسبات hasibat and حواسيب hawasib.

Due to this coordination ambiguity—i.e., the inability of WST in the study to recognize, for instance, the conjunction (و) waw and as a separate token (see Chapter Four, 4.3.1.6 Punctuation)—it was difficult to extract terms that include conjunctions such as (و) waw and, since terms that are preceded by conjunctions (head conjunction in Table 5.13) and compound terms that include conjunctions (term conjunctions in Table 5.13) share the same structure.

<table>
<thead>
<tr>
<th>Type of terms</th>
<th>English terms</th>
<th>Arabic equivalent</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>head conjunction</td>
<td>multimedia</td>
<td>وسائط متعددة</td>
<td>wasa'ī muta'dīda</td>
</tr>
<tr>
<td>term conjunction</td>
<td>parse</td>
<td>قراءة وتحليل سلسلة المحارف</td>
<td>qīrat wa tahlil sīlāl al maharīf</td>
</tr>
</tbody>
</table>

Table 5.13: Ambiguity involving Arabic coordination

---

42 One of the main sources of complexity of Arabic language morphology is the morpho-syntactic agreement between certain constructions such as noun and adjective and verb and subject. These agreement rules are more complex than simple matching rules found in languages such as English and French. These restrictions might involve discrepancy between morphological forms in specialized technical and scientific discourses.
The results depend largely on the tool used and whether it is easily adaptable to handle these specific features of Arabic script. Although most of the recently developed tools (including WST) have been developed to handle Arabic, it is not possible to process all the specific features of Arabic script. For instance, in the study corpora, the conjunction (و) waw and the bound phoneme (ة) ta marbuta are considered as separate tokens in some cases and parts of a token in others. (و) waw and (ة) is considered in Figure 5.3 as a part of the word ﻟﻜﻦ but (ولكن and but), whereas in Figure 5.4 it is a separate token. This phenomenon has been exhaustively discussed in the previous chapter (see Chapter Four).

5.2.4 Variation and Borrowed Terms

The simplest kind of influence that one language may exert on another is the borrowing of words, meaning that one language adds a new word or morpheme from another language to its
own lexicon. When there is cultural borrowing there is always the likelihood that the associated words may be borrowed too (Sapir 206), so it is not surprising that a highly productive term formation strategy of Arabic terminology is borrowing. Through these borrowings, according to Stetkevych (1970), Arabic is about to enter the Indo-European language family. Although Stetkevych goes too far, his claim is valid to certain extent. Borrowing from foreign languages is becoming one of the main ways through which Arabic is influenced, especially in a specialized arena such as computers and the Internet, thanks to new technology and international communication developments. This can be attributed to the fact that “translation activity in the Arab World is still too slow when compared to the speed with which new terms are coined for new concepts in the West and the frequency of their usage once coined” (Baker 88). It may also be attributed to the current linguistic situation in the AW and in the ANAC region in particular, where different languages are in contact (see Chapter Two).

Examination of the study corpora reveals that borrowing is the most frequent technique (among others such as literal translation) used for terminology creation. In C1, 45 borrowed terms have the highest frequency in the field of computers and the Internet. Some terms are not field-specific, for they can be found in other fields (e.g., kart 卡带 and kaïbl 卡插 cable), while others are specific to the domain of computers and the Internet (e.g., قرص بلو ری blue ray disc and ملتمیدیا multimedia multimedia). Table 5.14 shows the most frequently borrowed terms detected in C1, together with their English counterparts in C2.
Table 5.14: High-frequency borrowed terms in the two corpora

The values in Table 5.14 indicate that there is no great difference between C1 and C2 in terms of borrowing from foreign languages in general. The table shows that the term تكنولوجيا technology is the most frequent in C1 (231), whereas in C2 the term بروتوكول protocol has the highest occurrence (168), followed by كمبيوتر computer, which has 185 occurrences in C1 and 151 occurrences in C2. The two borrowed terms سكّنر scanner and موس mouse have also similar frequencies in both corpora. While سكّنر occurred 67 times in C1 and 73 times in C2, موس occurred 55 times in C1 and 41 times in C2. The biggest difference occurred in the term كتروني electronic, which occurred 158 times in C1 but only 73 times in C2. This indicates that there is a tendency among Arabic writers in general to use borrowed terms although there are established Arabic equivalents for many of these foreign terms. For instance، تقنية taqniya for technology، جهاز حاسوب hasub for computer، ماسح ضوئي scanner and فأرة mouse. The frequency of occurrences of ماسح ضوئي scanner and فأرة mouse in Table 5.12، for instance، is much lower than the frequency of occurrences of the loans سكّنر scanner and موس mouse in Table 5.14.
Lexical variations may occur as a result of borrowing from different sources. Table 5.15 shows variations in three borrowed terms. In the first example, the term لابتوب burtabl laptop is a direct loan of the French term ‘portatif,’ which means laptop, while the term لابتوب labtub is a direct loan of the English term laptop. In the second example, the quality of the vowel (whether long vowel in كابل kaïbl cable or short vowel in كابل kabl) is determined by way of approximation rather than by Arabic phonological criteria. The foreign vowels in cable were trans-positioned in Arabic كابل and كابل kabl. The term network typology was also used in C1 as طبولوجيا الشبكة tbuluẖya aṣabaka without the vowel (و)، whereas طبولوجيا الشبكة tbuluẖya aṣabaka with the vowel (و) was used in C2. The reason for this variation is also related to the trans-position of the short and long vowels in both English and French. The في في tbuluẖya aṣabaka was inserted to represent the long vowel /ai/ in the original English term typology /ˈtaɪpɒlədʒi/, whereas the omission of that في في tbuluẖya aṣabaka was made to transpose the short vowel in the original French term typologie. In other words, writers tend to spell the loaned term in a recognizable approximation of its original source, either English or French.

<table>
<thead>
<tr>
<th>English term</th>
<th>C1</th>
<th>Transliteration</th>
<th>C2</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>laptop</td>
<td>بورتابل</td>
<td>burtabl</td>
<td>لابتوب</td>
<td>labtub</td>
</tr>
<tr>
<td>cable</td>
<td>كابل</td>
<td>kabl</td>
<td>كابل</td>
<td>kaïbl</td>
</tr>
<tr>
<td>network typology</td>
<td>طبولوجيا الشبكة</td>
<td>tbuluẖya aṣabaka</td>
<td>طبولوجيا الشبكة</td>
<td>tbuluẖya aṣabaka</td>
</tr>
</tbody>
</table>

Table 5.15: Borrowing from different sources

5.2.4.1 Phonological Variations in Borrowed Terms

As noted above, writers from the two parts of the ANAC region vary in their use of borrowed terms. The diversity in the occurrence of loans appears to take place on the
phonological level by maintaining the source language pronunciation in the borrowed term. Writers adopt loans according to whether their linguistic source is French or English. It seems that writers’ background knowledge of either French or English increases the approximation to the original French or English pronunciation of terms. The borrowed term ‘ويب’ web, for instance, takes different forms when used as a loan in the two corpora. It is ويب wab in C2 and ويب wïb in C1 with additional (ي) infixed. Examples that show the impact of foreign language phonology are introduced in the following table:

<table>
<thead>
<tr>
<th>English term</th>
<th>Meaning</th>
<th>Loan in C1</th>
<th>Transliteration</th>
<th>Freq.</th>
<th>Loan in C2</th>
<th>Transliteration</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>modem</td>
<td>مودم (مرسل رقمي)</td>
<td>مودم</td>
<td>mudim</td>
<td>15</td>
<td>مودم</td>
<td>mudm</td>
<td>30</td>
</tr>
<tr>
<td>web</td>
<td>شبكة</td>
<td>ويب</td>
<td>wib</td>
<td>15</td>
<td>ويب</td>
<td>wab</td>
<td>14</td>
</tr>
<tr>
<td>email</td>
<td>بريد الكتروني</td>
<td>إيميل</td>
<td>imal</td>
<td>11</td>
<td>إيميل</td>
<td>imayl</td>
<td>58</td>
</tr>
<tr>
<td>Facebook</td>
<td>فيسبوك</td>
<td>فاسبوك</td>
<td>fashuk</td>
<td>47</td>
<td>فيسبوك</td>
<td>faïsbuk</td>
<td>19</td>
</tr>
<tr>
<td>transistor</td>
<td>ترانزيستور</td>
<td>ترانزيستور</td>
<td>tranzïstur</td>
<td>6</td>
<td>ترانزيستر</td>
<td>tranzïstr</td>
<td>14</td>
</tr>
<tr>
<td>virus</td>
<td>جرثومة</td>
<td>فيروس</td>
<td>faïrus</td>
<td>40</td>
<td>فيروس</td>
<td>fiyrus</td>
<td>45</td>
</tr>
</tbody>
</table>

Table 5.16: Loans from different linguistic sources

Table 5.16 above shows the spelling variations for six terms. For the first, the spelling مودم مودم was used 15 times in C1, and مودم مودم was used 30 times in C2. For the second, ويب ويب wïb web was used 15 times in C2, whereas ويب wab was used 14 times in C2. For the third term email, إيميل إيميل imayl was used many more times in C2 than إيميل إيمال imal in C1—58 times for the former and only 11 times for the latter. This is because إيميل إيميل imayl is more popular throughout the AW and commonly used than إيميل إيمال imal. For the last term, virus, the spelling فيروس فيروس faïrus was used 40 times in C1, while the spelling فيروس فيروس fiyrus was used 45 times in C2, showing that both spellings were used with similar frequencies in the two corpora. This means that the two spelling variations have similar popularity in the ANAC region. Two
important points should be considered here. First, some of the terms have an Arabic equivalent, as indicated in the second column from left (meaning). For example, *web* has the Arabic equivalent ِشَبَكَة ِṣhabaka, yet the French-derived word ِوِبَ wîb is widespread in C1 and the English-derived term ِوِبَ web is used in C2. Borrowed terms such as ِفِيروُسَ فيروسِ ِfaîrus virus, ِإِمِيلَ ِimayl email and ِمُودِمَ ِmudîm modem are also used even though they have Arabic equivalents (ِجرَثومة ِẖrthuma ِvirus, ِبريدَ اِلكترونيَي بريد الالكترونيِّ ِbarïd alktruny email, and ِمرسلْ رقْمْيَ ِmursl raqamy modem respectively). Secondly, some of these terms do not have an Arabic equivalent, and the borrowed word is the only existing word in the language. ِتَرَانْزِيْسْتُروَ ترانز١٠سُtor transistor in Table 5.16 is an example. Note that in the case of the term ِفِيروُسَ فيروسِ ِfaîrus virus, the sound ِ/v/ which does not exist in the Arabic phonological system, is often substituted by its nearest homorganic Arabic sound ِ(ِقَلَفَةَ ِقُلَفَةَ ِ/f/) in order to assimilate foreign sounds to Arabic.

However, various factors might lead to differences in loan terms. The pronunciation of these terms depends largely on:

1. The influence of a dialect, whereby dialects tend to adopt wholesale borrowed terms without any assimilation of their source sounds.

2. Level of education. On the one hand, people who have background knowledge of foreign languages may tend to imitate the original source of the term and pronounce it with phonological feedback from their foreign-language knowledge. On the other hand, people who have limited foreign-language knowledge will find it difficult to pronounce terms that include unfamiliar foreign sounds.

Finally, it should be pointed out that borrowing is a neglected area of study (Sager 14). Sager’s claim is still valid as far as Arabic specialized arenas, such as computers and the Internet, are concerned. While many studies have explored the phenomenon of borrowing into general
Arabic, few have been devoted to this phenomenon of borrowing into Arabic for LSP. Examples are Saraireh (2001) on inconsistency in new borrowed technical terminology into Arabic and Al-Qinal (2002) on the adoption of technical and scientific terms for which Arabic equivalents are unavailable. The field of computers and the Internet is also unusual, because it straddles special and general language in a way that differs from most other fields, such as chemistry and physics. Thus the use of computer terms is far-reaching and not limited to merely one social or professional group (see 3.7.1 for a detailed discussion about the features of the field of computers and the Internet). A more thorough and precise comparison, therefore, is needed to account for types of loans from foreign languages into Arabic in the domain of computers and the Internet. It would be of great interest for future research to examine this phenomenon in other specialized contexts.

5.2.5 Variation in Abbreviations

A very common term variation phenomenon in the domain of computers and the Internet is the usage of abbreviations, which are probably more common in English than in Arabic. Abbreviations are prevalent in both specialized and non-specialized domains and occur in both formal and informal contexts. They serve to shorten the linguistic code, thus allowing for more efficient and economic communication for terms and concepts that occur frequently in the discourse of a particular domain. Thus abbreviations are mostly introduced for terms longer than two words. However, some are introduced for one-word terms, such as ﺍﻟﻼإداري allaïdary non-administrative.

Abbreviated forms most often manifest themselves as initialisms and acronyms. While an initialism is formed from the initial letters in a compound term, like PC for personal computer and CD for compact disc, an acronym is a term formed from the initial letters and is pronounced
as a word, such as RAM for ‘random access memory’. Many abbreviated forms are more common than their full forms. For instance, the full form of HTML (Hypertext Markup Language) could be considered the marked form in the domain of computers and the Internet, as the HTML abbreviation is much more common than its full form. In the case of RAM (Random Access Memory), the full form is so infrequent that many users do not realize that RAM is an abbreviation, thus it becomes a new independent term. The domain of computers and the Internet offers a wealth of abbreviations. Table 5.17 shows some well-known examples of abbreviated terms in this domain:
Table 5.17: Some common abbreviations in the domain of computers and the Internet

Few methods for acronym and abbreviation formation in Arabic have been developed until recently, a fact that has attracted considerable attention (Spasic el al. 245), especially in scientific and technical domains. Thus there are no strict rules for creating and defining acronyms and abbreviations. In Arabic, as noted above, abbreviation is not as common as in English, since the language structure does not encourage abbreviation (Abu-Absi, 345). In English, abbreviation usually consists of the first letters taken from every word in the term (see Table 5.17 above). If we apply this method to Arabic, it would be difficult to form an acceptable abbreviation due to the cursive nature of Arabic script (words in Arabic cannot be divided, unlike
English) as well as the complicated sound system and derivation. The results would be as in the following table:

<table>
<thead>
<tr>
<th>English abbrv.</th>
<th>Source</th>
<th>Arabic abbrv.</th>
<th>Source</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTP</td>
<td>hypertext transfer protocol</td>
<td>ل ن ن</td>
<td>نا ق ل النص</td>
<td>lam nun nun</td>
</tr>
<tr>
<td>VPN</td>
<td>virtual private network</td>
<td>ش أ خ</td>
<td>ش آ خ ش آ خ كثوفة قاولية</td>
<td>sin alf kha</td>
</tr>
<tr>
<td>CPU</td>
<td>central processing unit</td>
<td>و م م</td>
<td>و م م و م م كثوفة قاولية</td>
<td>waw mim mim</td>
</tr>
<tr>
<td>HTML</td>
<td>hypertext markup language</td>
<td>ل إن ت</td>
<td>ل ن ن ن إلع ند النص الشبيبي</td>
<td>lam alif nun ta</td>
</tr>
</tbody>
</table>

Table 5.18: Complex Arabic abbreviation

If we want to read one of the Arabic abbreviations in the above table—for example, (ل ن ن) lam nun nun ‘HTTP,’ which is an abbreviation for لغة نقل النص ‘lughat naql alnas hypertext transfer protocol’—the reading would be ‘lam nun nun,’ because the structure for pronouncing one Arabic letter consists of more than three sounds (Lam CVC, noon CVC and non CVC). Thus it is difficult to form abbreviations depending on these sound conditions. For that reason, abbreviations are not commonly used in Arabic and are not usually accepted by Arabic linguists. Consequently, many terminological abbreviations in the two corpora are in fact English or French abbreviations that have been taken over into Arabic as is without applying any formation strategy.
Despite these constraints, some abbreviations find their way into Arabic and are now commonly used. They have been transliterated (such as كب taken from kilobyte), i.e., their phonology has been adapted to reflect the phonological system of the Arabic language.

The tendency in the study corpora is either to use the full form of the term without applying the phenomenon of abbreviation or to use the foreign abbreviation in Latin characters without rendering the form in Arabic script, as can be seen in Figure 5.5. The latter strategy would pose difficulties in understanding the meaning of the abbreviation for users who are not domain specialists and thus are not familiar with that foreign-language abbreviation.

![Figure 5.5: The use of foreign-language abbreviations in Arabic texts](image)

As can be seen in Figure 5.5, English abbreviations such as CPU Central Processing Unit, RAM Random Access Memory and MB RAM Megabyte of Random Access Memory were used in their original alphabet, without any attempt to find and use their Arabic equivalents or even translate the concepts into Arabic.

Statistics in Table 5.19 indicate that rendering foreign abbreviations into their complete Arabic translated forms is an important strategy in Arabic term formation. The table shows that terms such ashypertext transfer protocol, Hypertext Markup Language, central processing unit and virtual private network were found in C1 and C2 in their full Arabic equivalents with various
frequencies (13, 39, 25 and 45 respectively), immediately followed by their foreign language abbreviations.

<table>
<thead>
<tr>
<th>Arabic full term</th>
<th>English abbr.</th>
<th>C1 freq.</th>
<th>Texts</th>
<th>C2 freq.</th>
<th>Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>بروتوكول نقل النص التشعبي</td>
<td>HTTP</td>
<td>13</td>
<td>3</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>شبكة خاصة</td>
<td>VPN</td>
<td>45</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>وحدة المعالجة المركزية</td>
<td>CPU</td>
<td>25</td>
<td>3</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>لغة توصيف النص التشعبي</td>
<td>HTML</td>
<td>39</td>
<td>2</td>
<td>33</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 5.19: Use of full forms of abbreviations in Arabic texts

The use of the full term in Arabic, however, makes multi-word terms look like ordinary words. The use of capital letters in English abbreviated terms, such as HTTP, HTML, CPU and VPN, indicates the special status of these forms as scientific or technical multi-word abbreviated terms. This distinctive feature disappears in Arabic due to the fact that there is no capitalization in Arabic. For an exhaustive discussion about the issue of capitalization in Arabic see 4.3.1.5 (Capitalization) in the previous chapter.

Because of this and the complexity of using Arabic abbreviations (discussed above), writers regularly use foreign language abbreviations instead of their full-form equivalents in Arabic. The data indicates that the use of foreign abbreviations is greater than that of full forms in the two corpora. Table 5.20 shows the foreign abbreviated terms detected in C1 and C2.
A comparison of the results in Table 5.19 and Table 5.20 shows that there is a tendency in Arabic to use foreign abbreviations instead of their Arabic equivalents, either abbreviated or the full forms. The statistics in the two tables indicate that the abbreviation HTML, which has the highest frequency in the two corpora, was used 68 times in C1 in 3 texts and 57 times in C2 in 7 texts, whereas its full Arabic equivalent لغة توصيف النص التشغيلي was used 39 times in 2 texts. Similarly, the abbreviation CPU was used 47 times in 7 texts in C1 and 16 times in 3 texts in C2, while its full Arabic equivalent وحدة المعالجة المركزية was used 25 times in 3 texts only. The higher frequency of the foreign abbreviations in Table 5.20 indicates that although full forms of Arabic abbreviations are regularly used for certain abbreviations for which an Arabic equivalent is available, as in the case in Table 5.19, foreign abbreviations are still preferred. This analysis recalls the distinction made in Chapter Four (4.2.4 Corpus Studies and Terminology) between general and idiosyncratic usage of terms. In other words, certain abbreviations, such as CD and CPU, are used idiosyncratically by particular authors in a few texts, whereas others, such as

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
<td>68</td>
<td>3</td>
<td>57</td>
<td>7</td>
</tr>
<tr>
<td>CD</td>
<td>Compact Disk</td>
<td>38</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>RAM</td>
<td>Random Access Memory</td>
<td>33</td>
<td>4</td>
<td>39</td>
<td>6</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
<td>31</td>
<td>6</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
<td>47</td>
<td>7</td>
<td>16</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.20: Foreign language abbreviations
HTML, HTTP and RAM, have been generally accepted as abbreviations by a wider number of speakers and writers for the field of computers and the Internet.

The emergence of foreign-language abbreviations in Arabic comes as a response to the linguistic necessity imposed by technical development (Abu-Absi 341), especially if we know that Arabisation efforts carried out by academia in this regard seem to be very slow and not dynamic enough to keep up with the rapid advances in information technology. This usage also comes as a result of foreign-language hegemony, particularly English and French. In the corpora under study, foreign-language influence is reflected in the choice of abbreviations. Foreign-language preference, either English or French, appears to influence what abbreviation is used and in what lexical form. One prominent example of this phenomenon is the term information technology and communication, which in English is abbreviated as (ITC). The abbreviation ITC is used throughout C2, whereas the abbreviation TIC (technologies de l'information et de la communication) is used throughout C1. Similarly, the abbreviation ATM (Automatic Teller Machine) was used extensively in C2, while DAB (Distributeur Automatique de Billets) was used in C1, which could be incomprehensible and a source of confusion to users from other parts of the Arab World.

One particular problem that emerged in the analysis is that a single foreign-language abbreviation may account for different concepts. For instance, the abbreviation COM is used to refer to the following semantic references:

1. COM command file ملف أمر
2. .COM commercial domain نطاق تجاري
3. COM component مكون
COM is also used as part of a compound abbreviation, as in COMTI, which refers to ‘component object model transaction integrator.’ However, this problem is not unique to Arabic per se. In English, and perhaps many other languages, a single abbreviated term may account for different concepts, so that the meaning becomes very dependent on the context.

5.2.6 Terminological Variation and Neologisms

Neologisms are new words created in a language as a result of innovative development in fields such as science and technology\(^\text{43}\), including terms used in computers and the Internet. According to Baker (190), neologisms enjoy much acceptance in Arabic—for example, *software* برمجيات, *hardware* أجهزة.*

Although Arabic texts comprise the two corpora used in this study, the distribution of foreign terms represented in Latin script, mainly French, seems to be high. This, according to Wilmsen (193), is a common practice amongst writers of Arabic in treating novel concepts coming from outside of the AW for which no equivalent terms exist (see Figure 5.6). Examination of C1 reveals a perhaps excessive use of French terminology and the neglect of Arabic equivalents, even though the French term has been given an equivalent in Arabic, as in the case of the terms *informatique* المعلومة, *modem* مودم, *logiciel du navigation* برنامج, ‘مودم’ and *middleware* *برمجة* ملاحة.’ in Figure 5.6. The Arabic equivalents seem to be largely ignored by many writers. Furthermore, the word *mobile* is used extensively throughout the corpus despite the existence of four equivalents in Arabic (الآلة، التسجيل، المحمول، الخالوي).

\(^\text{43}\) There are many causes for creating new words in general language. Words, for instance are created just because people want to give a cool image or simply they like to play with their language, but technical and scientific innovative developments seem to be the leading factor behind the creation of new terms.
In many cases, the foreign term is used together with its Arabic equivalent as an additional reference point for the informed reader who might not be familiar with the Arabic term. Consider the following screenshot (Figure 5.7), where English terms such as favorites (misspelled as foavotites), refresh and scroll bar are used together with their Arabic equivalents (قائمة المفضلة qaïmat al-mufadhala, أعدة التحميل إعداد التحميل عادة التحميل ï'adat a tahmïl and شريط التمرير ñarît atamrîr respectively):

Figure 5.6: Excessive use of French in Arabic texts

Figure 5.7: Foreign terms used with their Arabic equivalents
Table 5.21 presents more examples of novel terms in the domain of computers and the Internet. The examples are all foreign concepts imported and spelled in the two corpora in their original Latin script.

<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
<th>C1 Freq.</th>
<th>C2 Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet forum</td>
<td>An online discussion where people can hold conversations in the form of posted messages.</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>Ethernet cable</td>
<td>Any of several types of coaxial cable used in Ethernets (dictionary.com)</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>DHTML</td>
<td>Dynamic hypertext markup language</td>
<td>16</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5.21: Terms used in their original Latin script

Table 5.21 shows that all the provided examples occurred with significantly similar frequency. *Internet forum* occurred 26 times in C2 and 17 times in C2, *Ethernet cable* has the same frequency in C1 and C2 (14 occurrences in each), whereas DHTML occurred 16 times in C1 and 28 times in C2. An explanation for this is that speakers of Arabic in general and in the ANAC region in particular are not conscious users of MSA or native speakers of the language (see Chapter Two, MSA and dialects). Arabic speakers are more familiar with English terms as a function of seeing English terms constantly on their screens or in documentation. They use interfaces that are partly or entirely still using foreign terms, mainly English. This would contribute to the excessive use of foreign terms instead of using existing terms in Arabic or creating new terms. The use of both Arabic terms and their foreign-language counterparts occurs as a result of the writer’s concern that the use of Arabic terms alone would cause a failure of communication.
In fact, Moroccans, Algerians and Tunisians often use French in their verbal and written communication instead of Arabic. Most online resources – not only computer and Internet resources, but also scientific and technical resources in general – are written in French all over the region. This is one of the main factors that influence the quality and quantity of Arabic term creation and the translation of scientific and technical terminology.

The attitude toward foreign languages, particularly French and English, could explain this increased use of foreign terms in Arabic in specialized areas such as computers and the Internet (See Chapter Two ‘language attitude’ for further discussion). The use and development of a language depends largely on the attitude of its speakers, which, in turn, depends on their interest in that language, their social rank and their educational level. This is seen, according to Jamai, as a conscious or even unconscious signal by an ‘interlocutor’ that she or he would like to suggest that they are educated and probably that they are westernized, equating the use of the language with open-mindedness and sophistication (51).

This is no surprise, since Arabic (MSA) is perceived by its speakers as a difficult language to learn (see Jamai 2008). In fact, its grammar and style are rather complex in comparison with those of other languages such as English and French. In a study of language attitude in Morocco, Jamai found that 43.8% of his respondents felt less comfortable with the use of Arabic language (MSA), compared with 56.2% who felt comfortable with the use of French.

On the other hand, it has been demonstrated in Chapter Two that French has very little influence on Arabic in Egypt and Libya. The influence of English on Arabic technical language in general is considerable in this part of the Arab World due to historical factors that have played a major role in term formation (see Chapter Two). Therefore, many English terms have become a
part of life in this region, as they are used dozens of times every day, not only as a matter of prestige but also because they have become an indispensable element in both professional and personal life. Examples include MP3, MP4, flash memory, and fax. Many of these terms have an equivalent adopted and accepted by many Arabic linguistic institutions (e.g., رقية الذاكرة السريعة raqiqa al-dakīra asarīa’a instead of فلاش ميموري flas mimury flash memory and البريد المصور albarid almusawar instead of فاكس faks fax). Nevertheless, people use these foreign terms not as a consequence of a need to refer to a new concept, but rather, in many cases, to make their language sound more modern.

Mohammed Al-Assal, a distinguished contemporary Egyptian linguist, warned at ‘The reality of Arabic Language in the Arab World’ conference, which was held in Cairo in 2008, that this increasing use of foreign terms leads to the marginalization of Arabic and threatens our culture.

إن زيادة عدد الألفاظ الأجنبية يهمل اللغة العربية ويشوهها ويعتبر أثر خطرًا على الثقافة داعيا إلى التعرف على مصادر هذه الألفاظ ومعانيها الأجنبية والبحث عن مرادفات عربية لها (13).

The increasing number of foreign words marginalizes Arabic language and distorts and threatens Arabic culture. We should identify the sources of these words and their meanings and then search for their equivalents in Arabic.

(Translation by the author)

Ennaji and Sadiqi have a different point of view. They think that “this type of bilingualism also serves the purpose of Arabization; the language contact is beneficial for the modernization of Standard Arabic, which borrows massively from French” (103). Ennaji and Sadiqi’s claim is true to a certain extent. Arabic has been in contact with other languages and, as a result, borrowing
from other languages, but mainly from English and French, into its lexicon is inevitable. It constitutes a strong method for lexical development of Arabic despite the cited resistance and despite warnings of marginalization and threats to Arabic language and culture. Many languages, including English and French, enriched and have been enriching themselves in this way.

5.3 Summary and Conclusion

This chapter aimed at giving a good picture of the phenomenon of Arabic terminology variation through comparison and analysis of the two corpora of computers and Internet texts generated from two Arabic regions. Terms extracted from these two corpora have been analyzed in order to provide answers to the research questions presented in Chapter One. The focus was primarily on orthography (e.g., spelling variations, vocalization, different loan translation, etc.), morphology (e.g., variations related to inflectional phenomena such as number and gender, derivational variations), lexicon (e.g., synonyms which are used interchangeably), phonological variations, and acronyms and abbreviations.

Ideally, terms of a special language discourse should be fixed items and should not be prone to any kind of variation. Terminologists, technical writers and translators have been trained to embrace terminological standardization and to disparage variation and employ consistency (Dury 2009, Bowker and Hawkins 2006). Term formation, therefore, has to be a conscious and well-thought-out activity (ibid) because the main purpose of terms is to facilitate specialized communication and knowledge transfer. However, despite this widespread assumption, the field of computers and the Internet in Arabic is one of the specialized areas in which variations are most apparent at all levels.

Despite the efforts of many Arabic institutions and organizations (see Chapter Six for more details about these efforts), whose aim has been to achieve terminological univocality in
the specialized domain of computers and the Internet (i.e., each concept is named by one term and one term refers to just one concept), this analysis has shown that term variation is present in the domain of computers and the Internet in the ANAC region. The analysis of the corpora has made it possible to infer patterns of variation as the linguistic source of loan terms. Writers adopt terms largely according to whether their linguistic source is French or English. Writers from Morocco, Algeria and Tunisia have a tendency to approximate the French pronunciation in loans, whereas writers from Egypt and Libya tend to approximate the pronunciation of original English terms. This has resulted in the existence of two different terminological references (French and English), which, in turn, has led to at least two Arabic terms for the same concept.

Because of the existing dialectical differences among countries of the ANAC region (discussed in Chapter Two), terminological variations in the domain of computers and the Internet were found among writers from different dialect areas. Writers and translators were apt to use terms in accordance with phonological features in their local dialects. However, insufficient evidence was found to prove this hypothesis due to factors such as the nature of the texts used in this analysis, which are collected from formal computer and Internet discourse. In fact, while non-standard Arabic dialects are largely used in informal discourses such as email communications and online chatting, standard Arabic is normally used in most formal online publications, including journals, magazines and online news.

Borrowing is the main method for term creation in the field of computers and the Internet. In fact, this field continues to adopt loans from different sources in order to cope with the new terminological needs. The development of new technology creates an ongoing need for new borrowings, and the creation of new computer and Internet terminology is certainly no
exception. Many of the concepts found in the corpus are represented solely by French and English terms that are orthographically modified to reflect foreign-language pronunciation. These terms find their legitimacy in Arabic because they do not have equivalents in Arabic or because of writers’ preference for using the foreign term instead of Arabic.

Finally, we may conclude that terminological variation is a significant phenomenon in technical languages (Jacquemin and Royaute 1994). Precise terminological descriptions are still needed, especially when it comes to Arabic terminology. Further analysis involving expanded corpus work remains to be done with regard to Arabic terminological variation, particularly in specialized fields such as computers and the Internet.
Chapter Six

Factors Affecting the Implementation of Successful Terminology Policy in the Arab World and

Conclusions
6.1 Summary of the Study

This dissertation is composed of six chapters: Chapter One has introduced the topic and has introduced the statement of the problem as well as the objectives of the study. It was hypothesized that two main factors, among others, determine Arabic term selection and use in the domain of computers and the Internet and thus lead to the phenomenon of term variation: 1) the influence and the hegemony of foreign languages, and 2) the influence of local dialects. Chapter Two reviews the introduction of Arabic into the region and its development from CA into MSA. In this regard, I have emphasized how the two concepts of Islamization and Arabization are related and how the linking of Arabic to the Qur’an affected linguistic practice as well as the development of Arabic terminology for centuries. Chapter Three shows that the top-down approach, whereby government agencies in charge of terminology planning do not listen to the practitioners in the field, is problematic. A successful LP and TP, however, needs a combination of top-down and bottom-up approaches (this is discussed further in this chapter). Chapter Four highlights some problems in achieving ease and reliability in carrying out even simple processing of electronic corpora in Arabic by means of automated tools. It is important, therefore, to extend the functionality of these tools to include the particular features of Arabic script, such as BiDi, complex morphology, and vocalization, in order to provide better support to users. In Chapter Five, the study investigates how foreign language hegemony have influenced the formation of the terms cited in the study. Although evidence to support the influence of foreign languages, mainly French and English, was found, not enough evidence was found to support the influence of regional dialects due to reasons already discussed in Chapter Five.
6.2 Introduction to Chapter Six

The current chapter is divided into six main sections. The first section contains a summary of the dissertation. In section two, I have identified six aspects that might explain the major factors that have impeded progress in creating and implementing an effective terminology policy with a view to standardizing Arabic computer and Internet terminology throughout the region. These include historical, sociolinguistic, political, economic, theoretical, and pedagogical aspects. No doubt, there are many other factors that must be taken into consideration, but it is beyond the scope of this study to describe them all due to the dynamic and flexible nature of terminology formation, which changes and evolves in accordance with different linguistic and cultural situations as well as different users and language needs.

The section treating the theoretical aspect includes a brief examination of the role and contribution of various agencies involved in the development of Arabic terminology. Although many agencies are involved in Arabic technical and scientific terminology production, the focus is on the most effective bodies in developing and promoting Arabic terminology. In this section, I also survey the existing term banks and examine what benefits the individual users and institutions might expect to derive from these term banks. A brief review of the challenges of pan-Arab coordination and cooperation in the field of technical terminology is also presented.

The chapter also proposes an alternative model for TP designed to ensure the involvement and coordination of different actors and stakeholders in TP implementation. The proposed model is not limited to the ANAC region or any particular environment.
involving terminology policy. The proposed model is general and is designed to be adaptable to any linguistic population. The importance of proper terminology training in the development of terminology activities will be highlighted, and the huge gap between terminology training at universities and the requirements of professional life will also be addressed. I conclude the chapter with a brief discussion of research limitations and suggestions of some topics for future research.

6.3 Factors Affecting Terminology Performance in the AW

There are many different viewpoints from which terminological variation can be explained (Freixa 52). I have already acknowledged that foreign-language hegemony and different regional dialects are not the sole factors that affect Arabic terminology choices. Many other factors should be taken into consideration. However, it is impossible to determine all these factors from this analysis due to the scope associated with this dissertation. In the following section, I will attempt to identify six aspects that might explain the major factors that have impeded progress in creating and implementing an effective TP with a view to standardizing and harmonizing Arabic computer and Internet terminology throughout the AW.

6.3.1 Historical Aspect

The study has already demonstrated that four languages are presently in conflict of in the ANAC region (i.e., Arabic, English, French and Tamazight). There is widespread official agreement that Arabic is the only politically, economically, and socially acceptable method of communication. At the same time, French and English are increasing in prominence as means for modern science and technology and for economic
development. There are also continued calls to use the Tamazight language, and its recent widespread use in radio broadcasts, magazines and newspapers is a result of the new democratic trend in the region. This conflict is historical in the sense that the ANAC region had been subject to different layers of colonialism, and of course these tensions are reflected in their linguistic manifestations.

In the historical survey of the introduction and development of Arabic in Chapter Two, I argued that the current problems that Arabic is having in developing its terminology are largely related to the ideology of Arabization and Islamization that has been imposed since the arrival of Arabs in the region. I have shown that the official LP is intended to impose Arabic language and Islamic identity and to deny the existence of multilingualism and linguistic variety, as well as to advocate a ‘monoglot ideology.’ The promotion of Tamazight language has never been carried out in a way that sincerely aims at creating a speech community in which this minority language has equal rights and uses with respect to Arabic.

Language, as Elzeer (212) puts it, should be allowed to follow its natural line of growth. It should not be pushed into directions that would most likely generate linguistic clashes. One may argue, however, this might not coordinate with the increased calls for LP and the implied desire to decrease the variability of terminological choices, especially in particular scientific and technical fields. Other languages in the region, such as Tamazight and colloquial dialects as well as foreign languages, might be taught in schools as a way to improve early childhood education and to increase the self-awareness of their speakers. They might also be allowed in universities, since MSA as a sole
medium of communication has failed to fulfil the needs of communication, particularly in specialized discourses such as computers and the Internet. More importantly, the implementation of the Arabization process as a LP should be disconnected from the concept of Islamization. The independence of the Arabization process will definitely allow the natural growth of MSA as well as the consideration of the natural influence of both local colloquial and foreign languages.

The collapse of the dictatorial regimes in Libya and Tunisia as well as the recent political and governmental reforms in Morocco and Algeria are leading to momentous economic and political reforms (or what is called ‘the Arab Spring’) which, among other things, are showing a loosening of the authorities’ grip over minority communities. Barring unanticipated shifts in the course of events, the geopolitical changes and the emergence of democratic trends in the region can be expected, at least in the long run, to provide the opportunity for Tamazight to be acknowledged as an unofficial sub-national language. Although these trends might endanger stability and unity and might not foster overall LP, they definitely encourage human rights and the preservation of minor languages. A realistic example is the former Soviet Union, where Russian remains a unifying lingua franca among the language communities of the old union and is still learned as the language of science, technology and commerce, but where other languages have gained regional rights under the new system. Furthermore, the democratic environments in many European countries protect the rights of minority social groups and give many minority languages in Europe official status. For instance, both Welsh in the

44 For instance, the current unstable situation in Libya and Egypt as well as the continuing tensions in Catalonia. The Basque had a long history of militant, and even terrorist, resistance to the central government, and the Catalans themselves are very much in the throes of an effort to separate from Spain.
United Kingdom and Irish in Ireland have protected language rights. Basque, Catalan and Galician have official status in Spain (Cenoz et al. 45).

Nowadays there is an increasing debate in the whole region revolving around the policy of Arabic as the only official language. The new democratic trend might also provide educational rights to the Amazigh population and might give practical support for Tamazight language and its implementation, e.g., teaching Tamazight in Amazighian educational settings as a second language. The immediate challenge in this regard is the shortage of implementation materials (school curricula, textbooks, assessment criteria, resource materials, etc.). Furthermore, the current language ideology connected to Arabic as the only language permitted in the educational system might also function as a major handicap in learning and teaching Tamazight.

Finally, this brief observation should be supported by more sociopolitical and sociolinguistic studies of language practice to reveal the multilayered tensions among these languages within a local or regional context. Other careful studies, therefore, are required with the aim of developing critical awareness of local and regional language practice of minority groups in order to raise overall language consciousness, thus expanding the standardization of MSA and reducing terminology variation. Such studies would provide language and terminology planners with data such as language preferences and attitudes and the ideologies of different language communities regarding LP and TP.

6.3.2 Sociolinguistic Aspect

Among the reasons for the current failure of LP and TP activities is the limited domains in which Arabic (MSA) is used. Although Arabic is the only official language in
all countries of the region, it is not the mother tongue of the Arabs, because it is taught at
schools almost as a second language (Ayari 250). “Unlike official language, mother
tongue serves in the daily life of the common person. People need not to go to school to
learn it and to know its codes…it begins to develop at the same time a child begins
breastfeeding and learns the first word with the utterance of that word by his mother.”
(Bereketeab 155).

Therefore, MSA is mainly used as a medium of communication in a narrow
demographic form only among educated Arabic speakers in conferences, gatherings of
Arabic scholars, and for formal school education. It is also used in printed media
(newspapers and magazines) and, to a certain extent, as a spoken medium in various
degrees in radio and television news. Spoken Arabic media, however, currently ranges
between the use of MSA and local dialects in many Arabic countries such as Egypt and
Lebanon in accordance with projected audience needs and preferences, affecting usage
and register. For ordinary conversation and non-formal dialogue shows, the colloquial
language is used, whereas for news reports and formal dialogue programs such as
political, religious and sports shows, MSA is used. This restriction in usage can be
attributed to the discrepancy between written and spoken Arabic forms (this phenomenon
has been exhaustively discussed in 2.4 ‘Arabic diglossia’). Even in formal school
education, MSA is used only in terms of reading and writing, as teachers often
communicate with their students and colleagues in local Arabic dialects.

Many Arabic scholars (e.g., Abu Saad 1987 and Ghazala 1977) have therefore
called for introducing dialect terms into the stream of MSA. Aub Saad (27), for instance,
appeals for the acceptance of local dialect terms in the standard context. He claims that
these terms enrich the standard Arabic vocabulary and fill in its lacunae, especially if these terms can be refined by returning them to their possible MSA origins. For him, this is better than contaminating ‘pure Arabic’ by allowing foreign terms and expressions to enter dictionaries. This is applicable to specialized domains, such as computers and the Internet, especially if we take into consideration the fact that computer and Internet terminology is gaining wider public use than ever before, thanks to the development of IT and the growth of Internet users. The acceptability of local dialect terms is justified in computers and the Internet domain because of its current dependency on wholesale borrowing from foreign languages in new term formation, taking into consideration the slow response of academia and terminology organizations in coining new terms in MSA. However, the question is: what kind of dialect should be given that status? And which dialect is the best, among others, to be standardized? Moreover, a dialect term assigned for a certain concept might vary from one Arabic country and/or region to another. What is really needed, instead, is to create a mechanism in the Arabic education system in schools and universities whereby MSA is the medium of communication so that people can internalize speaking and thinking in MSA naturally in addition to using local dialects, which in turn will reduce the adoption of foreign terms into Arabic. When speakers become more familiar with MSA, their need to use foreign terms that are incompatible with their mother tongue will decrease, allowing the adoption of features of MSA. The choice of MSA as the sole medium of instruction and communication might also lower the social recognition and prestige of English and French languages.
Part of the problem is the clearly unfavorable attitude that Arabic speakers have towards this language. I have demonstrated in 2.8 ‘Language Competition and Language Attitude’ that there is an ongoing shift in favor of foreign languages. This can be explained by the fact that Arabic (MSA) is perceived by its speakers as a difficult language to learn (see Jama 2008). This attitude is also because of the current archaic methods of teaching Arabic grammar (Darwish 2009).

As a result, large sectors of the Arab public do not command MSA sufficiently. This has made it difficult for them to develop successful scientific and technical terminologies and has forced even experts to use foreign terms for communication in scientific and technical domains. In fact, one of the likely reasons that the borrowed terms are used in preference to the standard terms (see the findings of data analysis in 5.2.4 ‘Variation and borrowed terms’) could be that writers and translators are not accustomed to using the standard Arabic terms or are simply unaware of them.

6.3.3 Political Aspect

This study has demonstrated that the complexity of the terminological situation in the ANAC is further exacerbated by political relationships. The process of terminology creation and standardization requires close collaboration amongst terminologists, subject specialists, linguists, translators and educators (UNESCO 16). In Chapter Three (3.8 ‘Terminology Practice and Official Terminology Policy in the AW’), I briefly demonstrated that the political situation in the region might be one of the potential causes for Arabic terminology variation. The fact that there are 22 countries with many political
issues among them poses a major obstacle to implementing a sustainable unified LP and to standardizing Arabic terminology.

Many decisions regarding LP and TP are initiated and formulated by government agencies, which hold the highest position on the administrative pyramid and dominate the economy. They lack the competence to make language policy or coordinate strategic language planning. They lack an understanding of even basic principles in, for instance, psychology and sociology, which could improve the quality of language education output. Policies tend to come directly from the authority, without any consideration for independent sources of expertise. Decisions are also largely underpinned by political government objectives. For instance, in Libya LP was used to achieve political stability and to prevent geopolitical conflict between the Arabs and Amazigh. For centuries, LP was more about ideology and the hegemony of Arabic than it was about pragmatic linguistic considerations. It gave privileged authority and legitimacy to Arabic and thus created a hierarchy in which other languages and cultures are neglected and marginalized.

This political factor seems to hamper the development not only of terminology infrastructure, but also general language infrastructure. In many cases, for instance, translators are still searching for terms in dictionaries instead of having online access to a comprehensive termbank\textsuperscript{45} like Termium or the Banque de Terminologie de Québec (BTQ). This can be largely attributed to the fact that Internet services are less developed and relatively expensive in the AW due to the restrictions that governments impose on communication technologies. A typical example is Libya: during the Gaddafi era a

\textsuperscript{45} Examples of Arabic term banks and terminological institutions are presented later in this chapter.
certain type of permission had to be obtained from the Libyan Interior Security Office in order to be able to obtain a computer, a fax machine, an internet connection and particularly a printer.

Paradoxically, another reason for Arabic terminological deficiency is the strong subconscious official resistance to foreign languages. In many countries politicians refuse to introduce English and French in elementary and even preparatory schools and insist that students should start learning (one of) these languages only after they are in complete command of MSA. For them, Arabic as the language of the Qur’an is beautiful and complete. In the case of Libya, for example, the official language policy was long used by policy makers as a mechanism to strongly resist the introduction of foreign languages such as English for merely political objectives. During the 1980s and 1990s, teaching of English was banned from Libyan schools and universities across the country as a result of political tension between Libya and the United States and United Kingdom (see Chapter Two for detailed discussion). Consequently, students start their translation classes with insufficient linguistic competence. Thus, their performance in translation and terminology is very poor; teachers end up correcting their grammatical mistakes rather than focusing on methodological and theoretical aspects. More discussion on terminology training problems in the AW and how to implement proper translation and terminology training will be highlighted below (6.2.6.Pedagogical Aspects).

6.3.4 Economic Aspect

Lack of sufficient financing for terminology programs and projects and lack of capital investment in infrastructure related to terminology, such as specialized software
tools (terminology extraction, terminology management, CAT, etc.) as well as “the huge price for starting up technical terminology in a newly computerized language” (DePalma and Joscelyne 30) seem to hamper the development of terminology in the region.

Terminology work in general is very expensive and time-consuming. The problem becomes even worse when it comes to building a term base or a term bank that involves Arabic, since a bilingual entry in Arabic and another language requires exhaustive human efforts in extracting terms, identifying equivalents, and verifying their use. It might also need extra hours of professional time due to the particularities and nature of Arabic script (discussed in detail in 4.3.1 ‘Problems in Corpus-building for Arabic’).

The major challenge is not to motivate those who have knowledge to contribute, but to motivate those who have the decision and power to participate, which is difficult. One of the greatest difficulties is to persuade and convince the stakeholders, particularly politicians, to allocate more money to terminology development and management. Furthermore, most terminology work is done in scientific and technical areas, and mostly at the request of the public sector. The private sector is not normally involved, and its interests are unwilling to pay for terminology projects. One main reason for this is the fact that the private sector does not play a major role in the language industry in general due to the long history of socialism in the region. Consequently, Arabic terminology work in general is still state-funded. Other interested bodies, such as national organizations and institutions, universities, specialized private organizations, publishers and media are less involved. The implementation of a successful terminology policy
requires significant investments in economies that can afford a substantial budget allocated to language and terminology matters.

6.3.5 Theoretical Aspect

The failure to implement a successful LP and TP in the AW can also be attributed to the poor theoretical background. There is little theory on Arabic TP (Siény 196). Terminology in general, compared with linguistics and many other disciplines, is a new discipline that has inadequate theoretical basis, which is considered by many theorists as problematic (see Cabré and Sager 1999, Sager and Nkwenti-Azeh 1990, Budin 2001, and Depecker 2002). Cabré and Sager, for instance, state that “so far existing analyses of terminology have not provided sufficient evidence for a coherent theory of terms.” (1)

The question here is not whether there is sufficient terminology theory, but rather whether the existing theory applies adequately to Arabic. There is no adequate Arabic theory for TP that can be particularly used in developing Arabic terminology practice. Elzeer (33) confirms that Arabic terminological thought is to a large extent in its infancy. This is basically related to the fact that terminology as a discipline is a relatively new field in the AW. Moreover, traditional methods in developing terminology have not been able to achieve satisfactory results or to meet the needs. It is therefore necessary to develop terminological theories that may provide frameworks that can be used to improve terminology practices. The integrative approach proposed later in this chapter (page 241) is one small tentative step in that direction. Although this approach, as well as the new insights about terminology variation in computer and internet contexts, could contribute to the development of new theories, methods and materials, more research on Arabic
terminology in general is needed in order to provide Arabic terminology planners with data based on elaborated theories and findings. Such a task is the responsibility of universities in particular, as well as institutions of terminology services (such as the Academy of the Arabic Language). The role of universities is to carry out research, to provide linguistic, methodological and technical consultation and instruction, to teach and train people about terminology and to evaluate the existing terminology work and methods.

The role of public institutions providing terminology services is to advise and support the government in the formulation, development, implementation and maintenance of terminological strategies and terminology development (UNESCO 16). The following section briefly examines the role and contribution of the main institution involved in the development of Arabic terminology, the Academy of the Arabic Language. This does not mean that terminology development in the AW is limited to this institution. In fact, it is an ad hoc task shared by many organizations and institutions, especially in technical and scientific domains, but it is beyond the scope of this study to examine all of them in detail.

6.3.5.1 Academy of the Arabic Language in Cairo

There are a number of official and unofficial institutions that also have the responsibility for Arabic language development and terminology production and standardization. Examples include the Academy of Arabic Language, the Bureau for the Coordination of Arabicisation in the Arab World (BCAAW) in Rabat (Morocco), the Arab Center for Arabization, Translation, Authorship and Publication (ACATAP) in
Damascus, and the Arab League Educational, Cultural and Scientific Organization (ALECSO). Nevertheless, most Arabic terminology work affecting theory and practice has been carried out by the Academy of the Arabic Language in Cairo. The Academy of the Arabic Language in Cairo46 (in Arabic مجمع اللغة العربية maẖma‘ al-lukha al-‘arabya) was founded in 1834 as The Royal Academy of the Arabic Language (المجمع الملكي للغة العربية al-maẖma‘ al-malaki lil-lukha al-‘arabya). The name changed to مجمع اللغة العربية maẖma‘ al-lukha al-‘arabya The Academy of the Arabic Language after the Jamal Nasir Revolution in 1952 and the fall of the Faruq kingdom. It is one of the most effective bodies in developing and promoting Arabic terminology. Its objectives are dedicated solely to the development of the Arabic language. According to the second article of the 1932 decree, the Cairo Academy’s objective is to:

make every effort to preserve the Arabic language and make sure it is able to meet the needs of contemporary science and art as well as those of modern life, by compiling dictionaries and similar works, and highlighting lexical and structural inaccuracies in Arabic usage, in addition to compiling a historical dictionary, and studying contemporary Arabic dialects in Egypt and abroad, and do all that is possible to ensure the progression of the Arabic language (cited in Elzeer 58).

46 Several other bodies are called the “Academy of Arabic Language.” For example, the Academy of Arabic Language in Damascus, Syria (the first one to be founded, in 1919); the Academy of Arabic Language in Iraq; the Academy of Arabic Language in Rabat (Morocco) and the Academy of Arabic language in Aman (Jordan), but the Academy of the Arabic Language in Cairo (Egypt) has the lead among them in developing Arabic terminology.
Most of its efforts in this regard are closer to publishing specialized dictionaries devoted to philosophy, linguistics and geology.—for example, its Philosophical Dictionary المعجم الفلسفى almuham Al-falsafy in 1982 and Dictionary of Narratology in 2002, Dictionary of the Terminology of the Koran in 1970. One of its most important dictionaries is the Arabic-Arabic dictionary المعجم الوسيط ‘al-maham alwasii, whose first edition was published in Cairo in 1962. The dictionary contains thirty thousand entries covering both CA and MSA, as well as six thousand illustrations.

The Academy started in 1834 with twenty committee members, and there are now sixty members from Egypt and different countries in the AW. According to Elzeer (43), they hold weekly meetings for the Egyptian members and an annual conference for members from other Arabic countries. At these gatherings terminological and Arabic language issues are discussed, information and experiences are exchanged, and fundamental decisions regarding general Arabic language and terminology are made. Thousands of new terms in science and technology are regularly defined according to Arabic patterns during these meetings and then combined in dictionary form. However, the terms created in this way are not always publicly accepted (see my discussion of acceptance of standardized terms in 6.2.5.3 ‘Standardization’). According to Wehr,

Relatively few of these have gained acceptance in common usage.

Specialists in all fields keep coining new terms that are either not understood by other specialists in the same field or are rejected in favor of others (viii).
Besides reasons discussed thoroughly in the study, such as language preference, language attitude, and MSA competence, this is because these terms are not created by subject-field experts and therefore their creators do not fully understand the requirements of the subject field. Instead, they are typically created by linguists who might not have specialized knowledge and who are also often neither interested nor trained in special subject field issues. This situation is further complicated by the fact that committee members are given no particular training on the principles and methods of terminology formation, and generally they learn from experience or from others, as many members have been in these committees for a long time. This is because they tend to be chosen for their age and prestige and thus are often out of touch with new trends in science, technology, and language itself.

There is no obligation on private and public institutions and individuals to use terms approved by these committees. Some scholars (e.g., Elzeer 165) therefore propose the idea of ‘enforcement acceptance.’ France, for instance, implemented similar enforcement against the use of English terms in an extreme manner (ibid). However, the academy has neither legal authorization nor the intention of enforcing terms it has created and/or standardized. Moreover, it might not be considered by many terminology users as an authority for planning and enforcing terminology. Intervention at a higher governmental level is needed because standards, according to Drame (31), are generally voluntary and become mandatory only when incorporated into national regulations or legislation. However, government intervention alone is not necessarily always effective. Terminology dissemination, therefore, should be done in a way whereby users of these terms are included in the process and a mechanism of interaction and collaboration
among specialized users and the committee on the one hand, and government institutions on the other is established.

6.3.5.2 Term Banks

A term bank (بنك مصطلحات in Arabic) is a collection, stored in a computer, of special language vocabularies, including nomenclatures, standardized terms and phrases, together with the information required for their identification. A distinction should be made here between term bank and term base. While the first is large in size and addresses a larger audience, e.g., companies and government and private agencies, the second is relatively smaller and addresses a limited audience. A term bank can be used as a mono- or multilingual dictionary for direct consultation, as a basis for dictionary production, as a control instrument for consistency of usage and term creation, and as an ancillary tool in information and documentation (Sager and McNaught 1). Term banks tend to be managed by public institutions with a significant administrative apparatus associated with the actual term base component of the term bank. They may also be associated with national language policies or terminology dissemination and implementation.

Computer-based processing, storing and dissemination of terminologies has been gaining ground in various public and private institutions in recent decades in many parts of the world. Over the last few decades, a large number of term banks have been established around the world, providing translation equivalents for technical terms from one language to another using bilingual and/or multilingual databases and providing detailed definitions and other linguistic and semantic information. Examples include the
following: the term bank operated by the Office de la langue française in Québec, established primarily to maintain the authenticity of the French language in Quebec (McNaught 114); the United Nations Multilingual Terminology Database, which provides terminology in the six official languages of the UN (Arabic, Chinese, English, French, Russian and Spanish); Termium Plus, the Federal Canadian Government terminology and linguistic database; and Multilex of Moscow State University. Some of these term banks have expanded their roles to serve a wider range of differing needs, such as the translation needs of organizations and standardization agencies. A detailed survey of the main international term banks is provided in McNaught (1988) and in Bennett, et al. (1986).

6.3.5.2.1 Arabic Term Banks

In the AW, there has been little development in the area of term banks. The number of established term banks is limited. One of the main reasons is the great deal of time and effort it takes to build up stocks of terminology, especially for languages that have no great terminological tradition or access to machine-readable terminological data, such as Arabic (McNaught 116). The oldest Arabic term bank is Ma’raby Bank (🤖 بنك معربي) LEXAR (المعجم العربي al-muḥam al-‘araby), which was started in 1960 and developed by the Institute of Studies and Research for Arabisation (Institut d'études et de recherches pour l'arabisation; IERA) in Rabat to promote the adaptation of Arabic to the economic, technical and scientific needs of the modern world (Sieny 175). Its aim is to electronically store and manage Arabic lexicographical data on the basis of the bilingual texts available at IERA. The bank has over 800,000 term pairs analyzed and stored in a
computer in three languages (Arabic, French and Latin) for more than 120 domains (insurance, banking, computing, etc.).

Another major project is BASM البنك الآلي السعودي للمصطلحات al-bank alaly as’udy lilmuš‘alalahat (the Saudi Terminology Data Bank), a multilingual termbase established in 1983 by the Saudi National Centre for Science and Technology (SANCST) in Riyadh. The project is still under development, and it currently has two hundred thousand terms, covering four major languages (Arabic, English, French and German). The terms are categorized into 21 domains and dozens of sub-domains, like information technology, law and agriculture. However, its interface is presently only in Arabic (see Figure 6:1).

Figure 6.1: BASM homepage

Other term banks include the بنك قم (قاعدة المعطيات المصطلحة) Terminology Database (Qimam) in Tunisia and مكتب تنسيق التعريب Arabization.org in Morocco. These term banks
are considered outdated because they depend on old published references. According to Daoud and Boitet,

Arabic term banks with its notably low informational coverage and outdated entries, for example (Arabization.org 2010), offers terminology service (considered one of the best term bases in Arabic language). However its specialized terminology depends on compiling old terms, for an instance in the domain of mathematics and astronomy altogether, the term base depends on a reference published in 1990, and has only 4067 entries, since then many terms has been introduced in those domains. (30)

Many of these term banks act only as electronic dictionaries and lexical databases, which means that their data are available as an online service that can be used for free, but cannot be accessed through the Internet and can be manipulated automatically only by a limited number of people who have authorized access, usually linguists and committee members. They are also of little real benefit to the majority of people who wish to use them, as they mostly deal with terminology as a list of alphabetically ordered lexical units. Figure 6:2 shows an example from BASM.
Building bilingual or multilingual terminological databases is not only a process of compiling lists of lexical items. It is a complicated process requiring significant reference information, such as domain identification, original contexts, definitions, and other resources. Recording all this information is quite difficult, because it needs not only technical and linguistic skills but also knowledge of Arabic morphology as well as of the specific domain. Moreover, not all of the entries are actually multilingual, as there are some terms missing in some of the languages. For instance, Figure 6.2 shows that in 12 multilingual entries 10 monolingual terms are missing. In terms numbers 305056 hybrid computer and 364772 computer graphics the Arabic equivalents are missing. Five terms (482711 \textit{computer to phone}, 454930 \textit{computer hardware}, 482188 \textit{computer output to laser disk}, 486845 \textit{computer ethics} and 219062 \textit{computer storage device}) are missing the French equivalents. Only in 111990 \textit{computer link} and...
In spite of these discrepancies, there is some cooperation among these term banks and international terminological institutions, especially in the area of scientific and technical terminology. For instance, the BAR project involved identifying existing terms and creating new terms in three languages (English, French and Spanish) in the domain of telecommunications and providing Arabic equivalents for these terms (Sieny 176). Another example is the cooperation between King Abdul Aziz City for Science and Technology (KACST) and Siemens, in which Siemens provided 25,000 terms in English, French and German in technical fields such as microelectronics and power engineering, while KACST provides their Arabic equivalents through its own term bank, BASM (ibid 175). It is painfully evident that each terminological institution and term bank in each Arab country has its own users, structure and objectives due to problems discussed throughout the study (see in particular the sociolinguistic and political aspects in this chapter). Darwish supports this claim:

Term banks and terminology networks remain strictly regional under the monopoly of certain agencies and the concept of sharing terminological data has not yet matured in the Arab World. In fact, it is sometimes undermined by the tendency to hoard terminologies for the sole purpose of accumulating terminological data competition. A number of Arabic lexicographers and terminologists store terminologies they have produced over two or three decades without making their work public (157).
Another major handicap in the development of terminology is the absence of interaction between various terminology users and these database resources with regard to the development of term entries. This approach is very traditional and top-down. It is also inconsistent with ISO 15188, which stresses that terminology products should be evaluated by terminologists, subject-field experts and users. Different means of user interaction such as feedback, suggestions or error reports are particularly important to measure and thus develop terminology quality. A typical example of such interaction is Terminologicentrum TNA (the term planning organization for Swedish), where terms are informally evaluated by users who are invited to comment on particular terms and send in their feedback. The feedback is then recorded, and if necessary term entries are corrected accordingly.

Obviously, countries and language communities are increasingly feeling the need to formulate systematic terminology policies (comprising also terminology cooperation strategies) in order to improve their competitiveness (UNESCO ii). What is really needed here is the establishment of a terminology organization that could act as an interface among agencies and organizations that are concerned with or involved in terminology as well as users working throughout the AW. Many Arabic scholars (e.g., Al-Kasimi 1979 and Sieny 1987) call for a central terminology body, bank or database. When he suggested the establishment of the Saudi Term bank BASM in 1983, Sieny (196) stressed the importance of coordination with other Arabic terminological work. For him, the dependence of different agencies and individuals on one central agency provided a practical basis for both coordination and cooperation among Arabic terminology
communities, which is especially important for eliminating duplication of effort and the creation of unnecessary and undesirable synonyms. In the AW, however, where there are 22 countries spread across two continents and a geographical area of more than 13 million square kilometers, it is very important to decentralize but also integrate terminology work so that national terminology institutions can act as references for both national and regional terminology organizations and agencies. These institutions can document terms in central or distributed, but related, databases and provide access to a unified terminological output for various subject fields. In practical terms, a pan-Arabic terminology association could be established with the goal of creating a cloud space where people (ordinary scholars, terminologists, translators, and other linguists) could share and update terminology. In general a decentralized TP would first consider the particular sociolinguistic and sociopolitical features of each Arabic country. These features, as it has been demonstrated in this study, differ not only from one country to another but also from one ethnic group to another. Second, it would allow for the implementation of an integrated LP and TP work throughout the AW rather than the hierarchic top-to-bottom through government institutions.

6.3.5.3 Standardization

ISO 15188 defines terminology standardization as “the establishment of terminology standards or of terminology selections in technical standards, and their approval by an authoritative body” (2). One of the causes of the complexity of Arabic terminology is the lack of standardization which, as has been already demonstrated

---

46 This sort of networking is known among database experts as ‘a network of aggregated resources’.
throughout this study, springs from the influence of foreign languages as well as varieties of Arabic language and, most importantly, the ill-defined understanding of terminology standardization. While considerable progress has been made in recent decades towards the standardization and unification of Arabic terminology, many technical and scientific loan terms, as indicated in the data analysis in Chapter Five, are used differently and have different meanings for different users, although they refer to one concept. Because of their frequent use, these terms are mistakenly considered by many users, even specialists in various technical fields, to be standardized terms that pose no communication problems.

As already noted in 3.8 (Terminology Practice and Official Terminology Policy in the AW), a clear distinction should be made between horizontal and vertical standardization. Horizontal standardization is defined as the standardizing of procedures or principles of terminology work, whereas vertical standardization is the standardizing of terminology for a specific subject field. Horizontal standardization is established by experts in terminology science, including linguists, language planners, terminologists, professional translators, and so forth, whereas vertical standardization is the product of subject specialists in a given field.

Ideally, terminology standardization should proceed in both dimensions, vertical and horizontal. UNESCO (18) emphasizes that the standardization of terminology should be based on standardized principles and methods and on scientific theory. There is misunderstanding in the AW of what terminology standardization might mean. For most who are currently involved in Arabic terminology work, terminology standardization
only involves vertical standardization, i.e., standardizing certain types of terms, such as those that have a variety in spelling or morphology, or those over which there is controversy. The emphasis is on inventing and documenting terms. TP and standardization, however, might be pursued with an emphasis on processes, not merely inventories of terms, and on management as well as the generation of terminologies (Hübschmannová and Neustupný 85).

Wehr and Cowan (10) observe that the problem of Arabic terminology lies not so much in inventing terms as it does in assuring that they gain acceptance. Many standardized terms are not accepted by user communities, including translators, proofreaders, localizers, educators, journalists, editors, public and private administrators, linguists, private and public companies and specialists working in bilingual contexts. Arabic media play a significant role in influencing the acceptance or non-acceptance of many new terms. For instance, a newly introduced term used by a creative newspaper writer or by a news broadcaster may gain currency and quick acceptance among the general language community, while linguists and linguistic bodies may prefer the standard designation of that term, if there is one.

In order to avoid terminological variation and eventually to promote standardization and thus dissemination and acceptance of terms, we need to address the issue of Arabic terminology standardization at four levels. Figure 6.3 is an illustration.
1. National (country) level: Terminological discrepancies have been found among users within the same country due to reasons mentioned throughout this study, such as different Arabic dialects and different language sources.

2. Regional level: There might be a need to standardize terms at the level of groups of Arabic countries with common sociolinguistic, historical and geographical circumstances, such as the Arabic North African countries (ANAC), the Middle East, and the Arabic Peninsula.

3. AW level: Unification and standardization of terms is desirable at the level of Arab countries in order to improve communication among companies working in different regions. The problem, however, is that many of the companies active in the area are French and German in the North African region of the AW and American and British in the Arab Peninsula and Middle East (not to mention the Chinese and South Korean companies working all over the AW). Therefore, they do not necessarily have a vested common interest in terminology standardization.
4. International level: It is extremely important to create a common scientific and technical language between consumers in the AW and their international suppliers, especially since these terms come from outside of the AW through new imports of concepts and products. Arabic standardization agencies should ideally work collaboratively with international organizations and standardizing bodies, such as the International Organization for Standardization (ISO), International Electro-technical Commission (IEC) and International Telecommunication Union (ITU). Currently, according to Wright (personal communication October 2013), there are no Arabic representatives in ISO TC 3747 and only Morocco and Tunisia are observing members.

This integrated terminology standardization, which involves close connection and communication among standardization agencies and users, would promote awareness of terminology issues and ideally might result in increased popularity and acceptance of standardized terms among Arabic and international communities. What is really needed is a concerted effort on the part of Arabic academics and intellectuals to lobby governments to support communication and participation activities with international organizations such as ISO.

A positive language attitude towards the standardized terms should be one of the core processes in this integrative standardization. This kind of effort should be pursued in a way that includes users of these terms. National and global users, experts and media

---

47 ISO TC 37 is a technical committee in the International Organization for Standardization (ISO) that prepares standards and documents concerning methodology and principles for terminology and language resources.
could be consulted before a standardized term is launched or a standardization recommendation is made. Elzeer (156) suggests that a very detailed investigation should be carried out into definitions, contexts and usage for the potential standard terms. For her, the work of unprofessional bodies such as ‘media-generated terminology’ should be disregarded because of the lack of transparent general policy and clearly defined rules and because they are discredited by linguists and linguistic bodies (57). Here approach, however, is the traditional top-down, as she ignores the significant role of term users. The integrative model proposed in this study emphasizes the involvement of the prominent figures in making Arabic terminology choices such as Arabic media. Moreover, the Academy of the Arabic Language in Cairo itself has approved many terms that had been generated by the media or even by an unknown party.

6.3.6 Pedagogical Aspect

Obviously, no progress can be achieved in terminology development without proper terminology training. Without proper training in terminology formulation and in techniques and procedures for establishing appropriate equivalents by finding the best equivalent that already exists or by using intelligent approaches for creating new ones, more anomalies are likely to be created by untrained and under-skilled translators (Darwish 74). The current discrepancies in Arabic terminology cannot be attributed only to the dichotomy of the foreign language reference (English and French), but also, and to a great extent, to the absence of proper training (Bhreathnach 2011).
Few academic institutions offer terminology training in the Arab World. Terminology training in the AW is available only at universities\(^{48}\), and only translation students are targeted for terminology training, mostly in the form of courses incorporated within translation programs, as in the translation program at King Saud University in Saudi Arabia, or as courses incorporated within English language departments, as in the translation program at Yarmouk University in Jordan. Many of these translation programs are focused on practical training (e.g., legal translation, commercial translation, financial translation), or what is called LSP training, rather than theoretical and methodological training. However, theoretical, methodological, and technical aspects must be covered in the training of terminologists (Cabré 222). Methodological training may cover topics such as term choice, language registers and term resources.

These programs also remain largely focused on linguistic issues, since the level of English skills of students is generally known to be poor. The reasons behind this deficiency have been extensively discussed in 1.6 with regard to the role and status of English in Libya. Linguistic competence, however, is only one of the many components that are expected of translation and thus terminology trainees. The interdisciplinary nature of terminology requires that Arabic trainees use relevant contributions from other disciplines. This expansion in Arabic terminology training would have a strong impact on the development of Arabic terminology work. Typically, a training course might include “elements from linguistics, particularly lexicology and lexical semantics, logic and classification theory, special subject fields, documentation, sociolinguistics, pragmatics

\(^{48}\) Associations, such as ATA (American Translators Association) in the US and its counterparts in Europe and Canada, offer trainings and workshops for translators, translation students and even terminologists.
and computing in addition to all knowledge related to a specific social situation” (Cabré 222).

Universities in the AW in general are suffering from a lack of sufficient references and books on translation and terminology theories. Most university students rely on their teachers to provide them with needed materials. This can be explained by the anti-westernization approach adopted by many Arabic countries such as Libya and Syria during the 1980s and 1990s (this phenomenon has been discussed in more detail in Chapter Three). Another important reason for this shortage of specialized translation and terminology materials is the fact that teachers themselves rely on foreign language textbooks and references because of the lack of locally published books in Arabic. A suggested example of a training manual could be Juan Sager’s textbook *A Practical Course in Terminology Processing*. Although the book was published by John Benjamins Publishing Company in 1990, it still provides an underlying discussion of major theoretical problems in the context of terminology practice. This includes cognitive, linguistic and communicative dimensions of concept theory. The book also deals with major problems involved in the compilation, storage, and retrieval of terminology.

More opportunities for training in terminology are also required, especially at advanced levels (doctoral or post-doctoral level), where the emphasis is particularly on methodology and theories rather than just on practice. However, preparing students for the real world, especially at advanced and professional levels, requires different kinds of training. The type of training depends largely on the type of task those students are going to carry out. It also depends on their own experience in terminology work in general and
in the specific tasks that will be assigned to them in real life. According to Cabré (20), a complete training program includes both training of terminologists who will take charge of systematic work and training of other specialists who assist in this work (e.g. scientists and technicians, translators, interpreters, technical writers, teachers, specialists in documentation, computational linguists, lexicographers), as well as the training of teachers of terminology.

Since translation and terminology training “is now almost entirely dependent on information technology” (Bowker 117), students enrolled in translation programs no longer work with the written text only. They need to master a wide variety of technical competences. First and most basically, they need sufficient background knowledge of standard Microsoft Office software such as Word processing and Excel, as well as Internet and email, which are prerequisite skills for translation and terminology students. This includes the ability to use some kind of search tools and familiarity with different research strategies such as Google search techniques. Second, and more precisely, students need the ability to use information management (IM), terminology management (TM), machine translation (MT) and computer-aided translation tools (CAT). Familiarity with these tools is becoming a prerequisite and of great importance not only for the work of today’s professional translators, but also for terminology trainees. They should be able to evaluate these tools in the sense of, for instance, what are the reliable tools for use in their language pairs? Which is the best to use for Arabic? Such evaluations may result in developments in the functionality of these tools with regard to Arabic-script-specific features as well as development of terminology pedagogy in general. In response to this
increasing “technicalisation” (Vienne 2000, cited in Schäffner, 99), any program aiming at professional translators and terminologists should include special IM, TM, MT, and CAT tools courses within its components, supplemented with other instrumental skills that can be useful, such as documentation and desktop publishing (DTP). These skills should be accompanied by what Gouadec called “information mining competence” (332): the ability to spot relevant resources and obtain any information required by different means and in various media. Vienne goes further and states that students “have to have the ability to evaluate and exploit the resources necessary to carry out the assignment received” (113).

The best training institutions, as Fatani (9) puts it, are the ones that provide students with access to a great deal of technology (professional-style translators’ work stations, software packages, specialist teachers, maintenance crews, etc.), since students’ future employment (in services, firms or organizations) and indeed survival as freelancers or in-house translators is closely linked to technology. One of the serious impediments to the progress of Arabic terminology is the widespread negative attitude towards CAT tools, not only among students, but also among professional translators. Although most Arabic translators already own tools that support terminological work and maintenance, they fail to harvest their many benefits. The tools tend to frighten off Arabic translators, especially freelancers, who might not be aware that these tools can lead to more job satisfaction since they can translate the tedious, repetitive material and allow human translators to tackle the more challenging translation work (Fatani 6). Among other reasons, this tools aversion can be attributed to the fact that Arabic translators’ level of computer skills is
generally known to be poor in comparison with, for instance, translators in United States or Europe. Recording data into a terminology management application, consequently, is quite difficult because it needs technical and computer knowledge, recalling the fact previously mentioned in this study that many CAT tools are not user-friendly when Arabic script is involved. Only a few universities have successfully integrated CAT tools into their mainstream coursework (Diana 2004, cited in Fatani).

The most commonly used options for both individuals and organizations in the AW are so-called ‘offline resources’, which include a simple paper card index, electronic dictionaries on CD-ROMs, print dictionaries, domain-specific reference books, or ‘online resources’, such as online dictionaries, glossaries, and machine translation such as Google Translate. In Saudi Arabia, for instance, “translators are still using a dictionary-based strategy, i.e., translations are done in a roughly word-by-word basis with no attempt to adopt a more communicative strategy or to address the Western audience of their translations” (Fatani 22). Consequently, as Darwish puts it, Arabic translators, regardless of their level of expertise, spend a great deal of their time in the translation process in a fruitless and often frustrating search for terms in available dictionaries, which are at best incomplete and inconsistent (74). This may also be attributed to the lack of computer literacy among Arabic translators. Many of them, according to Fatani (18), do not possess basic computer literacy and familiarity with PC computers and MS Windows environments. Translators insist that translation is “an ‘art’ which cannot possibly be reconciled with the machine.” (Fatani ibid).
Finally, it is necessary to note that the above-mentioned competences and skills are just a few of the most important and basic requirements. There are, undoubtedly, many other specific competences that should be taken into account in any terminology training program depending on what the trainees are going to do at the end of their program. It is also worth mentioning that designing a training program to master these specific competences depends largely on the context in which trainees are working, such as social and industrial needs and the predicted circumstances under which the future translator, terminologists, technical writers, proofreaders, localizers, editors and educators will be working. Elzeer (167) distinguishes four aspects of training that are related to terminology:

- Training of terminologists
- Training of specialists and others advising terminologists
- Training of professionals working closely with terminology
- Training/education of the general public.

Each of these scenarios involves a series of skills that need to be mastered on the part of the trainees as well as considered when designing a training program.

Obviously, there is an urgent need for Arabic translation and terminology trainees to be able to meet the existing technical challenges that they will inevitably face in the real world. Training institutions have a responsibility to equip trainees with these skills in order to enhance their employability. It is important that these technological competences be acquired and developed as early as possible in any terminology-training program.
Thus, an alternative terminology training approach is needed, but it needs to be implemented in a way whereby users are included in the process. The integrative approach proposed in this study might be considered in the sense that the design of training programs might be carried out in:

- Cooperation with interested bodies (such as national and international institutions, standardizing bodies and subject specialists). For instance, in cooperation with Infoterm, King Saud University (KSU) in Saudi Arabia is currently considering a training program in English that is intended to complement BASM (Siény 177).

- Consultation with terminology users (such as public and private organizations, publishers, professional translators, language practitioners and media). For example, a number of national and international institutions could be consulted and recommendations made before starting any terminology training program.

Involvement of users could be also useful in raising the importance of terminology training among them.

6.4 Conclusions

Nowadays there is a huge production of Arabic terms in various domains (Daoud et al 119 estimate: 50 terms daily, 17,500 yearly). Most of these terms are produced in scientific and technical domains, thanks to the current rapid development in science and technology, which in turn has confronted the Arabic language with the serious linguistic problem of expressing a wide range and increasing number of new concepts for which no
term exists in Arabic. Sager (6) speculates that “in developing countries, the natural
evolution of languages has been particularly threatened by the rapid expansion of new
fields of knowledge as a result of science and especially technology transfer.”

As a result, the influence of languages of developed countries, mainly English and
French, on Arabic terminology has been very significant. As shown in Chapter Five,
many foreign elements have been involved in Arabic terminology formation and
selection. Trends such as modernization, westernization, and globalization that have been
gradually invading Arabic culture since the age of colonization brought in a vast number
of foreign terms. This influence has also been more evident in recent years thanks to the
development and widespread expansion of communications media as well as extensive
travel, trade and tourism. These outside pressures take the form of a strong foreign
language hegemony, which results in the tendency of writers and translators, especially in
technical fields such as computers and the Internet, simply to adopt terms from English
and French. Many of these terms appeared with considerable frequency in my analysis.
Many of them have been assimilated into the Arabic language and are considered by
specialists in various technical fields to be standardized terms and thus pose no particular
problems.

The analysis of my data has also revealed that borrowing is one of the main means
of terminology formation. Many terms have been taken over from English or French,
either modifying their usage in Arabic or keeping them as is (with appropriate or
regionally influenced transliteration). This wholesale borrowing, which might endanger
the aesthetic of the Arabic language, has inspired many linguistic and terminological
bodies such as Arabic Language Academy in Cairo to try to put restrictions on this practice, despite its great potential for Arabic terminological and lexical expansion. This restriction, however, has failed to end the practice because of the hegemonic power of foreign languages, mainly English as a global ‘lingua franca’ of science, technology, business and media. It is also the result of the common belief among linguists that gaining a new foreign language term is a legitimate process and is not considered as a weakness in Arabic. Elzeer, for instance, argues that

Borrowing does not constitute any threat to the Arabic language. It is deeply rooted in the legacy of the language, and does not endanger any of its features, whether grammatical, structural or phonetic. Its use, therefore, is healthy, appropriate and much more needed (71).

Elzeer’s position would support allowing the Arabic language to follow its natural line of growth and would consider foreign language borrowing a natural phenomenon as long as there is contact between MSA and its neighboring languages, including colloquial Arabic dialects.

The competition between English and French has created two terminological sources: French-oriented source in Morocco, Algeria, Tunisia and Mauritania; and English-oriented source in Libya, Egypt and almost all of the rest of the AW. These two foreign language oriented-regions are distinct not only linguistically, but also culturally, socially and economically as a consequence of a long history of British and French occupation. According to Darwish (41), “two linguistically and culturally distinct blocs emerged in terms of foreign language affiliation. The colonial polarization eventually led to the polarization and definition of the Arab World’s linguistic subordinate.”
The analysis indicated that there are times that two or more borrowed terms from different sources exist for a single concept and there is at least one standard Arabic term competing against them. This variation is not a random act of defiance or carelessness (Bowker 1), but rather based on factors such as foreign language influence and preference. A writer may consciously choose to use a term that reflects her/his English orientation, while another uses another term that reflects her/his French orientation, even though both terms refer to the same concept.

Terminological variation, therefore, may occur as a result of foreign language approximation and borrowing from these two different language sources, as in the case of most of the examples provided in Chapter Five. However, some examples provided, such as the substitution of the letter qaaf .quant for the letter alef 1 (see 5.2.2.2 ‘Lexical Variation and Dialectological Influences’ for more examples) support another assumption the research made that terminological variation might be attributed to differences among Arabic dialects (see 6.1.1 Sociolinguistic aspect).

This discussion of the power of competing languages leads to the suggestion that LP interventions are necessary on both the level of status planning (SP, i.e., the status of language) and corpus planning (CP, i.e., the vocabulary of language) (see 3.4 Language Planning Typology). Although interventions in SP are more urgent than interventions in CP to develop new vocabulary and technical discourse for Arabic, CP is also required at this point in order to develop MSA in general as well as to prevent specialized ‘domain loss’. According to Alwood and Hendrikse (2003), many European languages were developed through an influential focus on CP. As a result, these languages are able to function as means for communicating specialized information and knowledge. In
German, for instance, the evolution of modern technical language occurred back in the 19th century. Under CP, the Germans sought to cut down Latin and French loans and coin new words based on German roots (Wright, personal communication October 2013).

I would also maintain that LP has been used to impose Arabic language and Islamic identity. However, the implementation of a successful policy on languages in general and on terminology in particular can be achieved by respecting, on the one hand, the inner disposition of language to evolve independently of any ideological bias, and on the other, the inherently aesthetic values of languages. The perception of these values tends to be influenced by the political culture people grow up in. Problems hindering language and terminology development in the AW might be the result of this connection between the two concepts of Islamization and Arabization.

Such a unified policy can be achieved only by adapting a more integrative approach to LP and TP. In Chapter Three, I demonstrated that LP and TP activities are generally top-down in developing countries (including the ANAC region), which involves a process whereby a linguistic community is manipulated by policy makers, usually politicians who do not leave anything to individuals to decide. People with power and authority make language-related decisions, often with little or no consultation with the ultimate language learners or users (Kaplan and Baldouf 196). I also argued that an effective LP and TP needs a combination of top-down and bottom-up approaches rather than a strictly top-down or bottom-up process.

However, these hierarchical approaches of top-down might create a gap between users, translators and subject matter experts on the one hand, and decision-makers and other stakeholders on the other, as the former (who are the end-users of terminology
work) are not directly involved in the decision-making process. The process here is a ‘one-way relationship,’ as it is concerned with official policy formation by authorities in control of power (Fishman 79) with no feedback option from or between users.

Instead of a strictly top-down or bottom-up approach, future LP and TP in the AW should be derived from a more integrative model in which governmental and non-governmental organizations as well as individuals play an influential role in determining the basic principles of this policy. LP and TP constitute a multifaceted phenomenon that is basically controlled and constructed from above but that cannot be understood unless introduced and participated from below by the ordinary persons who are the subject of these actions by governments and their spokesmen and activists for national movements (Hobsbawm 1). Decisions influenced by these activities are determined by both individuals (linguists, researchers, teachers, etc.) and formal organizations and institutions (universities, schools, professional associations, printing and publishing houses, Islamic schools, etc.). They also require official intervention (governments and their representatives) in order to be effective. Therefore, LP and TP in the AW should be based on an interconnected multi-directional ‘integrative model’.

The model involves close connection of policy makers with terminology users and close interaction with leaders involved in making terminological choices, particularly subject-field specialists and journalists. This concept is consistent with a recommendation by UNESCO, which emphasizes the “involvement of administrators, politicians, linguists, the media and the prominent figures in society” (25). Subject specialists, academics, national and international organizations and language practitioners should
also be involved. A very important aspect in the application of this approach is the public’s attitude toward language. The results discussed in Chapter Five show that writers use foreign terms more than their standard MSA equivalents. This might mean that foreign terms are more preferred than their Arabic counterparts and that MSA standardized terms are not always publicly accepted. Some of the attitudes towards foreign languages were already identified through the discussion in Section 2.8 ‘Language Competition and Language Attitude’.

These contradictory attitudes towards Arabic on the one hand and English and French on the other have caused a conflict with Arabic language practice regarding how Arabic should interact with these foreign languages and to what extent this interaction should be allowed. It is crucial, therefore, to create an atmosphere of preferring Arabic (MSA) terminology rather than foreign or even dialectal terms, especially in scientific and technical domains where Arabic domain loss is problematic. Surveys of attitudes toward Arabic (dialects and MSA) and foreign languages, as well as questionnaires and census reports carried out by ministries of education, universities, researchers and companies could be useful at this point in order to develop real assumptions and decisions on LP and TP based on empirical evidence. The role of universities in this regard is important to provide scientific foundations (methodologies and theories) for terminology development and to critically evaluate these methods. Specifically, the first challenge universities could take on is to establish an evaluative approach to assess and clearly describe the existing linguistic and terminology practice. This will provide fine-grained information about how languages and terminology are used and for what
purpose, before considering how to implement particular policies. However, as noted above, there is little terminology research in universities in the AW (see theoretical aspects and pedagogical aspects above).

Finally, it remains to be said that foreign-language influence on Arabic terminology has not yet ended. This influence will increase with the growth of globalization. In the ANAC region in particular and the AW in general and with the tendency toward English and French as a function of modernization, there will be an even greater increase of foreign terms in MSA. Such an influence will continually foster new foreign terms that in many cases fill lacunae and even compete with their established Arabic equivalents. Sager (6) suggests that “the flow of terminology into a society with a foreign technology is greater than the filter can handle”. Arabic therefore has no option but to interact with the surrounding languages and gain from them. It will continue to follow this tradition of adopting and foreign terms that have been absorbed through linguistic, cultural, economic, technical and political interaction.

I recognize that there are many problems involved here, and many others that must also be addressed, because there is no perfect model for TP, since terminology is a dynamic discipline. Thousands of terms are produced annually, and a rigid approach will not solve the problem. However, the application of such an interconnected approach could contribute towards a rapid and more consistent growth of Arabic terminology. One cannot say how this approach can actually be implemented, particularly under the current political and economic crisis in the ANAC region and Middle East, but no doubt future research development can turn to answer this question as well as many others raised throughout this study.
6.5 Limitations and Future Research

This dissertation presents only a preliminary investigation of multilingualism in the ANAC region and its influence on term formation and use, with the hope that this investigation will lead to a better understanding of the motivation behind terminology variation in the field of computers and the Internet. Much work remains to be done in the field of terminological variation.

Firstly, the study is restricted to the examination of computers and the Internet terminology in Arabic. It would be interesting therefore to explore the occurrence of variation in other specialized technical and scientific domains. It would be also interesting to explore whether or not term variation occurs in Arabic regions other than the ANAC.

Secondly, the focus of the investigation was on the applicability of the major patterns of terminology formation (e.g., derivation, compounding and borrowing) put forward by Arabic scholars. However, technical terms can in general be formed in a great number of ways that tend to be applied differently to different subject fields and to different languages (Bowker 488). Furthermore, the range of examples discussed in Chapter Five does not represent an exhaustive list of all variations of computers and Internet terms, especially if we take into consideration the dynamic nature of this domain. It is likely that future research might find more applicable methods and patterns of Arabic terminology formation and more appropriate examples in the future.
Finally, corpus research is useful in gathering both conceptual and linguistic information as well as usage information about the terminological units (Sager 133), but it can prove to be limited by the challenge of reliable data collection and representation. Although every effort was made to find texts for the two corpora, there were limitations caused by the lack of relevant electronic Arabic resources on the WWW in general; especially for Arabic specialized texts, it was not easy to find a representative quantity of MSA texts that deal with computer and Internet contexts and that are available in electronic format. This is especially problematic when creating the corpus for Morocco, Algeria and Tunisia, where related materials are in French rather than in Arabic. Moreover, PhD research is usually restricted in scope and time. No claims can therefore be made about the generalization of the dissertation findings to all Arabic linguistic situations, particularly those in the Arab Peninsula and Middle East. The possibilities offered by corpus tool analysis of large quantities of linguistic data are far from mapped out (Bowker 494). Future research with adequate quantities of MSA data will definitely be more helpful in understanding what motivates the formation and use of Arabic terminology.

Whatever future research remains to be done, this study is significant in providing a strong basis for the future of Arabic terminology work by proposing an alternative approach that guarantees the involvement of different actors and stakeholders and thus further development of Arabic terminology work. The study is also significant in providing new insights by identifying and analyzing the types of difficulties that undermine terminology work in the AW. However, various external aspects have an
influence on the process of terminology formation and thus have to be considered in future research. Some of these aspects are dynamic and cannot be predicted, particularly in the current swift turn of events in the region.

**Appendix I: Websites from which corpora were collected.**

<table>
<thead>
<tr>
<th>Website/Magazine</th>
<th>Arabic Version</th>
<th>URL</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabicseo.net</td>
<td></td>
<td><a href="http://www.arabicseo.net/">http://www.arabicseo.net/</a></td>
<td>C2</td>
</tr>
<tr>
<td>Moroccan E. Governance</td>
<td>الإدارة الإلكترونية المغربية</td>
<td><a href="http://www.itp.net/arabic/">http://www.itp.net/arabic/</a></td>
<td>C1</td>
</tr>
<tr>
<td>Financial Technology in Algeria</td>
<td>التكنولوجيا المالية في الجزائر</td>
<td><a href="http://cte.univ-setif.dz/coursenligne/medchaib/index.html">http://cte.univ-setif.dz/coursenligne/medchaib/index.html</a></td>
<td>C1</td>
</tr>
<tr>
<td>World of Educational Technology</td>
<td>عالم التقنية التعليمية</td>
<td><a href="http://edu.tech-wd.com/">http://edu.tech-wd.com/</a></td>
<td>C2</td>
</tr>
<tr>
<td>The Technical Books Library</td>
<td>مكتبة الكتب التقنية</td>
<td><a href="http://www.kutub.info/library">http://www.kutub.info/library</a></td>
<td>C2</td>
</tr>
<tr>
<td>Agence Tunisienne d’Internet</td>
<td>الوكالة التونسية للانترنت</td>
<td><a href="http://www.ati.tn/ar/home.php">http://www.ati.tn/ar/home.php</a></td>
<td>C1</td>
</tr>
<tr>
<td>National Bureau of Education and Training ONEFD</td>
<td>الديوان الوطني للتعليم والتكوين</td>
<td><a href="http://www.onefd.edu.dz">http://www.onefd.edu.dz</a></td>
<td>C1</td>
</tr>
<tr>
<td>Edunet/Tunisia</td>
<td>البوابة التربوية التونسية</td>
<td><a href="http://www.edunet.tn/">http://www.edunet.tn/</a></td>
<td>C1</td>
</tr>
<tr>
<td>Morocco Soft Magazine</td>
<td>جامعة هواري بومدين للعلوم والتكنولوجيا</td>
<td><a href="http://www.usthb.dz/ar/spip.php?article73">http://www.usthb.dz/ar/spip.php?article73</a></td>
<td>C1</td>
</tr>
<tr>
<td>Hawary Bomadian for Science and Technology</td>
<td>الرقمية مجلة تكنولوجيا المعلومات</td>
<td><a href="http://mag.mpcn.biz/">http://mag.mpcn.biz/</a></td>
<td>C2</td>
</tr>
<tr>
<td>Digital Magazine for Information Technology</td>
<td>المجلة العربية للانترنت</td>
<td><a href="http://www.ati.tn/ar/home.php">http://www.ati.tn/ar/home.php</a></td>
<td>C1</td>
</tr>
<tr>
<td>Tunisian Agency for Internet</td>
<td></td>
<td><a href="http://www.mincom.tn/index.php?id=11&amp;L=1">http://www.mincom.tn/index.php?id=11&amp;L=1</a></td>
<td>C1</td>
</tr>
<tr>
<td>Ministry of Information and communication technology</td>
<td></td>
<td><a href="http://www.mpcn.biz/">http://www.mpcn.biz/</a></td>
<td>C2</td>
</tr>
<tr>
<td>Moroccan Magazine/Technology</td>
<td></td>
<td><a href="http://clyousfi.jeccan.com/categories/">http://clyousfi.jeccan.com/categories/</a></td>
<td>C1</td>
</tr>
<tr>
<td>Computer Magazine</td>
<td></td>
<td><a href="http://walhaseb.com/">http://walhaseb.com/</a></td>
<td>C2</td>
</tr>
</tbody>
</table>
### Appendix II: Stop List of Arabic Words and Their English Equivalents

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>أيها</td>
<td>here</td>
</tr>
<tr>
<td>آخر</td>
<td>last</td>
</tr>
<tr>
<td>إذا</td>
<td>if</td>
</tr>
<tr>
<td>أراد</td>
<td>wanted</td>
</tr>
<tr>
<td>أسوأ</td>
<td>worst</td>
</tr>
<tr>
<td>أفضل</td>
<td>best</td>
</tr>
<tr>
<td>إلى</td>
<td>to</td>
</tr>
<tr>
<td>الآن</td>
<td>now</td>
</tr>
<tr>
<td>أنت</td>
<td>you</td>
</tr>
<tr>
<td>أو</td>
<td>or</td>
</tr>
<tr>
<td>أي</td>
<td>any</td>
</tr>
<tr>
<td>أيضا</td>
<td>also</td>
</tr>
<tr>
<td>بدون</td>
<td>without</td>
</tr>
<tr>
<td>بعد</td>
<td>after</td>
</tr>
<tr>
<td>بين</td>
<td>between</td>
</tr>
<tr>
<td>تقريبا</td>
<td>almost</td>
</tr>
<tr>
<td>الذي</td>
<td>which</td>
</tr>
<tr>
<td>ثم</td>
<td>then</td>
</tr>
<tr>
<td>جيد</td>
<td>good</td>
</tr>
<tr>
<td>حتى الآن</td>
<td>yet</td>
</tr>
<tr>
<td>حصلت</td>
<td>got</td>
</tr>
<tr>
<td>حيث</td>
<td>where</td>
</tr>
<tr>
<td>الذي</td>
<td>which</td>
</tr>
<tr>
<td>الذين</td>
<td>who</td>
</tr>
<tr>
<td>رئيس</td>
<td>major</td>
</tr>
<tr>
<td>سوف</td>
<td>will</td>
</tr>
<tr>
<td>ضروري</td>
<td>necessary</td>
</tr>
<tr>
<td>via</td>
<td>probably</td>
</tr>
<tr>
<td>عندما</td>
<td>when</td>
</tr>
<tr>
<td>غالبا</td>
<td>often</td>
</tr>
<tr>
<td>فقط</td>
<td>only</td>
</tr>
<tr>
<td>في</td>
<td>in</td>
</tr>
<tr>
<td>في الأساس</td>
<td>basically</td>
</tr>
<tr>
<td>في حين</td>
<td>while</td>
</tr>
<tr>
<td>قد</td>
<td>may</td>
</tr>
<tr>
<td>كان</td>
<td>was</td>
</tr>
<tr>
<td>كانت</td>
<td>she was</td>
</tr>
<tr>
<td>كان</td>
<td>object</td>
</tr>
<tr>
<td>كثير</td>
<td>many</td>
</tr>
<tr>
<td>كل</td>
<td>each</td>
</tr>
<tr>
<td>كما</td>
<td>as</td>
</tr>
<tr>
<td>كيف</td>
<td>how</td>
</tr>
<tr>
<td>لا</td>
<td>no</td>
</tr>
<tr>
<td>لا يزال</td>
<td>still</td>
</tr>
<tr>
<td>لا يمكن</td>
<td>can not</td>
</tr>
<tr>
<td>لأن</td>
<td>because</td>
</tr>
<tr>
<td>لديك</td>
<td>you have</td>
</tr>
<tr>
<td>لديه</td>
<td>has</td>
</tr>
<tr>
<td>لكن</td>
<td>but</td>
</tr>
<tr>
<td>لنا</td>
<td>us</td>
</tr>
<tr>
<td>هن</td>
<td>his</td>
</tr>
<tr>
<td>لها</td>
<td>her</td>
</tr>
<tr>
<td>لهم</td>
<td>them</td>
</tr>
<tr>
<td>لي</td>
<td>me</td>
</tr>
<tr>
<td>ليس</td>
<td>not</td>
</tr>
<tr>
<td>ما</td>
<td>what</td>
</tr>
<tr>
<td>مثل</td>
<td>such as</td>
</tr>
<tr>
<td>مع</td>
<td>with</td>
</tr>
<tr>
<td>معظم</td>
<td>most</td>
</tr>
<tr>
<td>من</td>
<td>from</td>
</tr>
<tr>
<td>من قبل</td>
<td>by</td>
</tr>
<tr>
<td>نحن</td>
<td>we</td>
</tr>
<tr>
<td>هذا</td>
<td>this</td>
</tr>
<tr>
<td>هم</td>
<td>they</td>
</tr>
<tr>
<td>هنا</td>
<td>here</td>
</tr>
<tr>
<td>هناك</td>
<td>there</td>
</tr>
<tr>
<td>هو</td>
<td>he</td>
</tr>
<tr>
<td>هؤلاء</td>
<td>these</td>
</tr>
<tr>
<td>هي</td>
<td>she</td>
</tr>
<tr>
<td>و</td>
<td>and</td>
</tr>
<tr>
<td>وبالتالي</td>
<td>thus</td>
</tr>
<tr>
<td>وجود</td>
<td>presence</td>
</tr>
<tr>
<td>يريد</td>
<td>wants</td>
</tr>
<tr>
<td>يكون</td>
<td>be</td>
</tr>
</tbody>
</table>
### Appendix III: Corpus One Term List

<table>
<thead>
<tr>
<th>Arabic Term</th>
<th>Meaning</th>
<th>Freq.</th>
<th>Sect.</th>
</tr>
</thead>
<tbody>
<tr>
<td>شبكة الحاسوب</td>
<td>Network</td>
<td>476</td>
<td>10</td>
</tr>
<tr>
<td>الكمبيوتر</td>
<td>Computer</td>
<td>435</td>
<td>24</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet</td>
<td>241</td>
<td>70</td>
</tr>
<tr>
<td>تقنية</td>
<td>Technology</td>
<td>231</td>
<td>17</td>
</tr>
<tr>
<td>MD</td>
<td>Media</td>
<td>183</td>
<td>17</td>
</tr>
<tr>
<td>الكمبيوتر</td>
<td>Computer</td>
<td>183</td>
<td>10</td>
</tr>
<tr>
<td>Windows</td>
<td>Windows</td>
<td>149</td>
<td>19</td>
</tr>
<tr>
<td>البريد الإلكتروني</td>
<td>Email</td>
<td>140</td>
<td>14</td>
</tr>
<tr>
<td>الاتصالات العالمية</td>
<td>Worldwide</td>
<td>121</td>
<td>16</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet (v4)</td>
<td>84</td>
<td>12</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet (v6)</td>
<td>74</td>
<td>9</td>
</tr>
<tr>
<td>شبكة خاصة</td>
<td>VPN</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>البرمجيات</td>
<td>Software</td>
<td>42</td>
<td>10</td>
</tr>
<tr>
<td>ذاكرة الوصول العشوائي</td>
<td>RAM</td>
<td>97</td>
<td>12</td>
</tr>
<tr>
<td>Windows</td>
<td>Windows</td>
<td>87</td>
<td>15</td>
</tr>
<tr>
<td>HTML</td>
<td>HTML</td>
<td>68</td>
<td>12</td>
</tr>
<tr>
<td>فيسبوك</td>
<td>Facebook</td>
<td>47</td>
<td>7</td>
</tr>
<tr>
<td>ويندوز</td>
<td>Windows</td>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>فيسبوك</td>
<td>Facebook</td>
<td>47</td>
<td>4</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
<td>47</td>
<td>7</td>
</tr>
<tr>
<td>الكمبيوتر</td>
<td>Technology</td>
<td>46</td>
<td>12</td>
</tr>
<tr>
<td>الكمبيوتر</td>
<td>Computer</td>
<td>45</td>
<td>9</td>
</tr>
<tr>
<td>مضادات الفيروسات</td>
<td>Anti-viruses</td>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td>اللوحة الأم</td>
<td>Mother board</td>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td>لغة تطوير النص</td>
<td>HTML</td>
<td>59</td>
<td>7</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet</td>
<td>58</td>
<td>6</td>
</tr>
<tr>
<td>الشبكة الحكومية</td>
<td>Internet</td>
<td>58</td>
<td>6</td>
</tr>
<tr>
<td>CD</td>
<td>CD</td>
<td>38</td>
<td>5</td>
</tr>
<tr>
<td>بروتوكول</td>
<td>Protocol</td>
<td>36</td>
<td>16</td>
</tr>
<tr>
<td>لوحة الجمعية</td>
<td>Main board</td>
<td>36</td>
<td>15</td>
</tr>
<tr>
<td>شبكة الويب العالمية</td>
<td>Internet Web</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>RAM</td>
<td>RAM</td>
<td>33</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English Term</th>
<th>Meaning</th>
<th>Freq.</th>
<th>Sect.</th>
</tr>
</thead>
<tbody>
<tr>
<td>الحواسيب</td>
<td>computers</td>
<td>31</td>
<td>6</td>
</tr>
<tr>
<td>HTTP</td>
<td>HTTP</td>
<td>31</td>
<td>6</td>
</tr>
<tr>
<td>إكسپلورر</td>
<td>Explorer</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>الحاسوب</td>
<td>Computer</td>
<td>27</td>
<td>8</td>
</tr>
<tr>
<td>وحدة الإخراج</td>
<td>Output device</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td>كمبيوتر</td>
<td>Laptop</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet forum</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>وحدة المعالجة المركزية</td>
<td>CPU</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>إدغ</td>
<td>Edit</td>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>الفروست</td>
<td>Vирус</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>الحاسب</td>
<td>Computer</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>كلب</td>
<td>Dialogue box</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>Excel</td>
<td>Excel</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>شبكة للإنترنت</td>
<td>World Wide Web</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>مايكروسوفت</td>
<td>Microsoft</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>محاوي</td>
<td>Computers</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>ترجمة</td>
<td>Scanner</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>ترجمة</td>
<td>Project</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>HTML</td>
<td>HTML</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>وحدة التخزين</td>
<td>Power unit</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Internet</td>
<td>Internet</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>نظام التشغيل</td>
<td>Operating System</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>جهاز</td>
<td>Aja</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>المكتبة الرسومية</td>
<td>Web</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>حاسب</td>
<td>Computer</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>لوحة المفاتيح</td>
<td>Keyboard</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>كابلית الماسح</td>
<td>Ethernet cable</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>تصفح المواقع</td>
<td>Program Temp</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>الحاسوب</td>
<td>Laptop</td>
<td>13</td>
<td>7</td>
</tr>
</tbody>
</table>
## Appendix IV: Corpus Two Term list

<table>
<thead>
<tr>
<th>Arabic Term</th>
<th>Meaning</th>
<th>Freq</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>الإنترنت</td>
<td>Internet</td>
<td>239</td>
<td>32</td>
</tr>
<tr>
<td>البروتوكول</td>
<td>Protocol</td>
<td>160</td>
<td>30</td>
</tr>
<tr>
<td>الكمبيوتر</td>
<td>Computer</td>
<td>151</td>
<td>33</td>
</tr>
<tr>
<td>الشبكة</td>
<td>Network</td>
<td>137</td>
<td>28</td>
</tr>
<tr>
<td>شبكة الإنترنت</td>
<td>Internet</td>
<td>130</td>
<td>30</td>
</tr>
<tr>
<td>شريدة البيانات</td>
<td>Text bar</td>
<td>122</td>
<td>18</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet</td>
<td>105</td>
<td>7</td>
</tr>
<tr>
<td>حاسوب</td>
<td>Computer</td>
<td>99</td>
<td>10</td>
</tr>
<tr>
<td>الويب</td>
<td>Web</td>
<td>97</td>
<td>13</td>
</tr>
<tr>
<td>عربة</td>
<td>Dialog box</td>
<td>83</td>
<td>16</td>
</tr>
<tr>
<td>الحاسب الألي</td>
<td>Computer</td>
<td>79</td>
<td>15</td>
</tr>
<tr>
<td>الحواسيب</td>
<td>Computers</td>
<td>71</td>
<td>7</td>
</tr>
<tr>
<td>الألكتروني</td>
<td>Electronic</td>
<td>72</td>
<td>21</td>
</tr>
<tr>
<td>لوحة المفاتيح</td>
<td>Keyboard</td>
<td>72</td>
<td>8</td>
</tr>
<tr>
<td>مايكروسوفت</td>
<td>Microsoft</td>
<td>69</td>
<td>16</td>
</tr>
<tr>
<td>بروتوكول</td>
<td>Protocol</td>
<td>67</td>
<td>12</td>
</tr>
<tr>
<td>القرص الصلب</td>
<td>Hard Disk</td>
<td>60</td>
<td>5</td>
</tr>
<tr>
<td>القناة</td>
<td>Wire</td>
<td>44</td>
<td>17</td>
</tr>
<tr>
<td>ميورورفت</td>
<td>Microsoft</td>
<td>42</td>
<td>11</td>
</tr>
<tr>
<td>متصفح</td>
<td>Browser</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td>إيميل</td>
<td>Email</td>
<td>55</td>
<td>15</td>
</tr>
<tr>
<td>ويندوز</td>
<td>Windows</td>
<td>59</td>
<td>12</td>
</tr>
<tr>
<td>HTML</td>
<td>HTML</td>
<td>57</td>
<td>19</td>
</tr>
<tr>
<td>البريد الإلكتروني</td>
<td>Email</td>
<td>56</td>
<td>12</td>
</tr>
<tr>
<td>بطاقة الذاكرة</td>
<td>Memory card</td>
<td>55</td>
<td>12</td>
</tr>
<tr>
<td>البرمجيات</td>
<td>Software</td>
<td>52</td>
<td>13</td>
</tr>
<tr>
<td>ويندوز</td>
<td>Windows</td>
<td>51</td>
<td>30</td>
</tr>
<tr>
<td>الإنترنت</td>
<td>Internet</td>
<td>49</td>
<td>12</td>
</tr>
<tr>
<td>حاسوب</td>
<td>Computer</td>
<td>47</td>
<td>6</td>
</tr>
<tr>
<td>الاعضاء</td>
<td>Access</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>المعالج</td>
<td>Processor</td>
<td>43</td>
<td>19</td>
</tr>
<tr>
<td>الحاسب</td>
<td>The computer</td>
<td>42</td>
<td>29</td>
</tr>
<tr>
<td>الموس</td>
<td>Mouse</td>
<td>41</td>
<td>5</td>
</tr>
<tr>
<td>HTTP</td>
<td>HTT</td>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td>RAM</td>
<td>RAM</td>
<td>39</td>
<td>6</td>
</tr>
<tr>
<td>القرص الصلب</td>
<td>Floppy disk</td>
<td>36</td>
<td>15</td>
</tr>
<tr>
<td>الشاشة</td>
<td>Screen</td>
<td>36</td>
<td>5</td>
</tr>
<tr>
<td>لوحة القيادة</td>
<td>Main board</td>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>برامج</td>
<td>Software</td>
<td>33</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th>Language Description</th>
<th>33</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>العربية</td>
<td>Language description</td>
<td>33</td>
<td>12</td>
</tr>
</tbody>
</table>

[Data Table]


Al-Sulaiti, L., and E. Atwell. “Extending the Corpus of Contemporary Arabic.”


---. Designing and developing a corpus of contemporary Arabic (2004). Print.


Becker, V. *A transfer grammar of the verb structures of Modern Literary Arabic and Lebanese Colloquial Arabic.* Diss. Yale University, 1964. Print.


263


<http://www.arabicwata.com/WATA_Magazine/Articles/index.html>