BRINGING THE LEXICAL APPROACH TO TAFL: EVALUATING THE PRIMARY LEXICON IN PART ONE OF THE AL-KITAAB FII TA’ALLUM AL-’ARABIYYA ARABIC AS A FOREIGN LANGUAGE TEXTBOOK SERIES

by

Janelle Nicole Moser

A Thesis Submitted to the Faculty of the

The School of Middle Eastern and North African Studies

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF ARTS

In the Graduate College

THE UNIVERSITY OF ARIZONA

2013
STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

____________________
Janelle N. Moser
SIGNED: Janelle N. Moser

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Sonia Shiri
Sonja Shiri
Assistant Professor of Arabic Linguistics

____________________
04/26/2013
Date
ACKNOWLEDGEMENTS

I would like to thank a number of people who helped to bring this project to fruition. First I would like to thank my committee chair, Dr. Sonia Shiri, for her time, energy, and patience in guiding me through this process. Without her supervision and mentoring, completing this project would not have been possible. I would also like to thank my committee members Dr. Samira Farwaneh and Dr. Grace Fielder for their guidance and insight as well as providing me with the theoretical tools to make this project conceivable. I would be amiss if I did not thank Tim Buckwalter and Dilworth Parkinson for responding to my inquiries and allowing to me use their *Frequency Dictionary of Arabic* for this project. I am also thankful for the tireless support of my family, friends, and colleagues during the writing process, especially at times when completing this project seemed like an insurmountable feat.
# TABLE OF CONTENTS

**LIST OF FIGURES** ..................................................................................................................................................... 6  

**ABSTRACT** .................................................................................................................................................................. 7  

**INTRODUCTION** .......................................................................................................................................................... 8  

**CHAPTER 1: THE IMPACT OF DIGLOSSIA, ORTHOGRAPHY, MORPHOLOGY,**  
**AND SEMANTIC TRANSPARENCY ON THE ACQUISITION OF VOCABULARY**  
**AND READING ABILITIES IN MODERN STANDARD ARABIC (MSA)** ............................................................................ 14  
  *Diglossia* ........................................................................................................................................................................ 15  
  *Impact of diglossia on acquisition* ................................................................................................................................. 18  
  *Alphabet and Orthography: Challenges to processing* ................................................................................................. 20  
  *Root and pattern: the morphological base of MSA* ........................................................................................................... 21  
  *Derivational Morphology* ................................................................................................................................................ 22  
  *Inflectional Morphology* .................................................................................................................................................. 28  
  *Adjectives, number, agreement, and case* ....................................................................................................................... 30  
  *Other challenges: “degree of transparency” and degree of synonymy* ......................................................................... 31  

**CHAPTER 2: DEFINING WORDS, THE MENTAL LEXICON, AND THE LINK**  
**BETWEEN READING AND VOCABULARY** .................................................................................................................. 32  
  *What is a word: Units of counting* ................................................................................................................................. 32  
  *Defining the mental lexicon* .............................................................................................................................................. 36  
  *Stages of vocabulary acquisition and the role of the root and word* ........................................................................... 38  
  *Dividing the mental lexicon: aspects of word knowledge and the receptive/productive divide* ................................ 40  
  *Receptive Vocabulary: Size and importance to reading* ............................................................................................... 43  
  *Frequency and text coverage in English and Arabic* .................................................................................................... 44  
  *Corpora: Utility and SLA applications* ........................................................................................................................ 49  
  *Textbook Studies, Corpora, and the Evaluation of “Vocabulary Levels”* ................................................................. 49  

**CHAPTER 3: METHODOLOGY AND SUBJECTS** ......................................................................................................... 54  
  *Methodology* .................................................................................................................................................................. 54  
  *The textbook: The Al-Kitaab Fii Ta’allum al-‘Arabiyya, Part One* .............................................................................. 59  
    *Pedagogical philosophy and structure* ........................................................................................................................ 59  
    *Vocabulary* ................................................................................................................................................................. 60  
  *The frequency data: A Frequency Dictionary of Arabic* ............................................................................................... 64
CHAPTER 4: RESULTS, DISCUSSION, PEDAGOGICAL RECOMMENDATIONS, AND CONCLUSION

Part One: Minimal word families from Al-Kitaab fii Ta’allum Al-‘Arabiyya Part One……..67
Part Two: Evaluation of vocabulary based on frequency data ........................................70
1-999 Word Level ...........................................................................................................70
1000-1999 Word Level ................................................................................................71
2000-2999 Word Level ...............................................................................................72
3000-3999 and 4000-5000 Word Levels .....................................................................73
Combined totals ........................................................................................................74
Patterns of divergence between textbook vocabulary and corpora-derived data ........74
“Broken” Plurals .........................................................................................................75
Alternative and Additional Broken Plurals ...............................................................76
Divergent Broken Plurals and Singular Nouns ..........................................................77
Divergent verbs, verbal nouns, and other forms ......................................................78
Discussion of results .................................................................................................79
Pedagogical recommendations ..................................................................................82
Directions for future research and conclusion .........................................................85

APPENDIX A: MINIMAL WORD FAMILIES FROM AL-KITAAB PART ONE,
ALPHABETICAL ORDER ..............................................................................................88

APPENDIX B: TEXTBOOK VOCABULARY ARRANGED BY FREQUENCY ..........145

REFERENCES ..............................................................................................................162
LIST OF FIGURES AND TABLES

Figure 1.1-Lexical doublets from MSA and spoken Lebanese Arabic..................17
Figure 1.2-Chart of Arabic Verb Patterns.......................................................23
Figure 1.3-Verbal nouns for triliteral and quadriliteral roots..........................26
Figure 1.4-Deriving active and passive participles...........................................27
Figure 1.5-Derived nouns...............................................................................28
Figure 2.1- Additions to a word family at different levels of inflection and affixation........35
Figure 2.2-Word families for k-t-b..................................................................36
Figure 2.3-A sketch of word knowledge for qa:l (to say)..................................41
Figure 2.4- Conceptualizing Levels Based on Vocabulary (Van Mol: 2000)...........47
Figure 3.1-Example of “Learn These Words” section......................................61
Figure 3.2-Example of “Remember these words”...........................................62
Figure 3.3-Primary vocabulary list before introduction of derivational morphological processes........................................................................................................63
Figure 3.4-Limited word families in primary vocabulary after presentation of some derivational and inflectional morphology................................................................63
Figure 3.5-Text type and sources for written texts............................................64
Figure 4.1-Totals and means from Part One Word Families...............................68
Figure 4.2-Word family for k-t-b (write) from Al-Kitaab Part One.....................69
Figure 4.3-Coverage rates of textbook vocabulary, 1-999 word level..................71
Figure 4.4-Coverage rates of textbook vocabulary, 1000-1999 word level..........72
Figure 4.5-Coverage rates of textbook vocabulary, 2000-2999 word level..........73
Figure 4.6-Overall Coverage rates of textbook vocabulary, 1-5,000 word levels...74
Figure 4.7-Singular nouns with plurals unattested in the Frequency Dictionary...75
Figure 4.8-Alternative Broken Plurals.............................................................76
Figure 4.9-Additional Broken Plurals...............................................................76
Figure 4.10-Divergent singular/plural noun pairings........................................77
ABSTRACT

This study proposes two models for exploring the lexical contents of Part One of the most popular Arabic as a Foreign Language textbook series, *Al-Kitaab fii Ta’allum Al-‘Arabiyya*. Through the lens of a word families framework, this study hypothesizes on the contents and arrangement of the L2 Arabic mental lexicon after completing the textbook. Through counting lexemes, lemmas, and word family members, it is possible to gain insight into the quantity of vocabulary items present within the textbook outside of traditional measures like the triconsonantal root. Through a frequency-based framework, this study analyzes textbook vocabulary items in light of the 5,000 most frequent lemmas in the language from a corpus of 30 million tokens from *A Frequency Dictionary of Arabic* (Buckwalter and Parkinson: 2011). A comparison between textbook vocabulary and frequency data points to the relationship between the vocabulary studied by AFL learners and the most widely used forms in the language as a whole. While this study gives special consideration to frequency data up to the 3,000 word level, the sheer amount of lexical knowledge necessary for reading Arabic newspapers and novels necessitates integration of frequency-derived data at even the novice level. A lexical and frequency-based approach to AFL instruction and curriculum design may prove helpful in decreasing the decidedly large vocabulary burden (Nation: 1990, cited in Young: 2011) for learners of Arabic as a Foreign Language.
INTRODUCTION

The lexical or vocabulary-based approach to SLA (Nation: 2001; Nation and Meara: 2002; Nation and Waring: 1997; Schmitt: 2001) places the multi-faceted concept of vocabulary knowledge at the center of second and foreign language acquisition. Knowledge of the most frequent words in the language, derived from spoken and written corpora, has been utilized to estimate a “threshold” of vocabulary knowledge necessary for text coverage (Hu and Nation: 2000; Laufer: 1992; Nation and Waring: 1997; Nation: 2006) as well as a tool to evaluate vocabulary levels and contents of language textbooks and instructional materials in the context of English as a Foreign or Second Language (Biber, et al.: 2004; Chujo: 2004; Chujo and Genung: 2003; Chujo and Nishigaki: 2003; Korprowski: 2005; Meunier and Gouverneur: 2009; Nelson and West: 2000; Romer: 2004a, 2004b, 2006b), French as a Foreign Language (Waugh and Fonseca-Greber: 2002), Spanish as a Foreign Language (Godev: 2009), and German as a Foreign Language (Lipinski: 2010). While recent studies by Belkouche, et. al (2010) and Chik, et. al (2012) represent tentative steps into the general field of textbook studies in Arabic, neither specifically addresses the lexical contents of AFL textbooks or the relation of contents to frequency data. To date, there has been little effort within the TAFL context to use empirical data to analyze the contents of the most widely utilized materials or textbooks, studied by an estimated 35,000 students nationwide (Modern Language Association: 2010, p. 2). This study represents a first step in cataloguing and analyzing the quantity and quality of hypothetical receptive knowledge that can be obtained after completion of Part One of the popular TAFL textbook series, *Al-Kitaab fii Ta’allum Al-‘Arabiyya* (Brustad, et. al: 2004).
Chapter One presents an overview of the Arabic language in general, including a discussion of language-specific features that may impede rapid vocabulary acquisition by native speakers of spoken Arabic as well as learners of Arabic as a Foreign Language. Chapter Two includes a general presentation of the word knowledge framework in SLA. This section will also cover conceptualizing the overall contents of the “mental lexicon” (Van Patten and Benati: 2010) including word families and general concepts of interlanguage (Selinker: 1972) and multicompetence (Cook: 1992). It will also review the divide between receptive and productive knowledge, and the link between these types of knowledge and receptive activities such as reading. This chapter will also cover general uses of corpora within SLA as well as studies merely compiling corpora and wider applications to assessment of foreign and second language textbooks and materials.

Taking a page from textbook studies utilizing corpora-derived data as a framework of analysis, this study will favor a “page-by-page” approach (Romer: 2004a; Gilmore: 2004; Hyland: 1994; Koprowski: 2005; Godev: 2009; Lipinski: 2010) over a corpus-based approach. The former is preferred over the latter, in which the sum of all the language presented within textbooks is privileged over saliency of specific vocabulary items (Nelson and West: 2000; Biber, et. al: 2004; Meunier and Gouvernour: 2009). Despite the appeal of the “pedagogical corpus”-based approach, where analyzed data is extracted from a constructed textbook corpus based on frequency within the textbook (Meunier and Gouvernuer: 2009), the “page-by-page” approach brings focus to vocabulary that will most likely be taught or studied, given focus through direct or contextual activities throughout the
chapter, and therefore is the vocabulary most likely to be acquired (Stahl and Fairbanks: 1986).

As evidence points to the importance of the first 3,000 word families or approximately 5,000 discreet words as the “turning point” of receptive knowledge for reading (Laufer: 1997), with the first 1,000 word families or 1,600 words covering 77% of written texts (Nation: 2001), this study will ultimately examine what kind of minimal receptive vocabulary knowledge for reading could be obtained after completing Book One of the Al-Kitaab series. Part one of the study will propose minimal word families from the primary (“Words” section) and secondary (“Remember These Words”) vocabulary lists. Given evidence that may support the role of the consonantal root as well as the word as a representative unit in the Arabic FL mental lexicon (Ravid: 2005; Jiang: 2004) and the English language evidence proposing psychological reality as an adequate basis for word families (Nagy, et.al: 1989; Bertram, Laine & Virkkala, 2000, cited in Nation: 2012), this study will organize textbook vocabulary according to word families as a basis for describing the quantity of vocabulary within the textbook. Part one of the study will be guided by three primary questions:

1) How many total lexemes (including MWUs) could a student recognize receptively if only the primary vocabulary from Part One of the series was memorized?

2) How many lemmas (excluding MWUs) could a student recognize receptively if only the primary vocabulary from Part One of the series was memorized?

3) On average, how many family members do the minimal word families contain?
Hypothetical word families will be based on Bauer and Nation’s (1993) definition as a “base word and all of its derived forms that can be understood without having to learn each form separately” (p. 253). Initial counts for the total number of lexemes (Schmitt: 2000, p. 2) and lemmas will consider not only discrete words but also “multi-word units” as these items are more likely to be “learned, stored, and used as complete units” (Nation and Meara: 2002, p. 36). It will also include the arithmetic mean ($\bar{x}$) of the number of members per family in these hypothetical word families after Nation (1983, cited in Laufer: 1997, p. 24). This section will also include observations regarding observed patterns within word families.

Part two of the study will examine the quality of members of minimal word families in light of data from the *Frequency Dictionary of Arabic* (Buckwalter and Parkinson: 2011). This corpus-derived dictionary meets the first of Biber, et. al’s (1998) requirements for a corpus as a representative sample of natural language, as the dictionary was culled from a corpus of 30 million tokens of 90% written and 10% spoken texts (Buckwalter and Parkinson: 2011, p. 3). Buckwalter and Parkinson’s efforts to present final frequency counts adjusted for range across each text type within the corpus strengthens the *Dictionary* as a tool to examine the “actual patterns of language use” within modern Arabic (Reppen and Simpson: 2002, pp. 92-93). Part two of the study will be guided by the following questions:

1) What kinds of coverage rates of the 1,000-5,000 word levels within the *Dictionary* can be achieved using textbook vocabulary?

2) What kinds of patterns of divergence emerge when analyzing textbook vocabulary in light of frequency data?

Coverage rates will be assessed by calculating the number of lemmatized lexical items found at the 1,000, 2,000, 3,000, 4,000, and 5,000 word levels in the same vein as
previous textbook studies utilizing a word-levels approach to vocabulary analysis (Chujo and Nishigaki: 2003; Chujo and Genung: 2003; Chujo: 2004; Godev: 2009; Lipinski: 2010). Due to indications of the importance of the 3,000-3,500 word levels in Arabic (Van Mol: 2000; Al-Batal: 2006), analysis will be carried out for the first 1,000-3,000 words in 100-word increments. Overall coverage rates at each level will also be compared. Analysis of reoccurring patterns of divergence between the two data sets will also be carried out to inform actionable pedagogical recommendations outside simply targeting word level increments or levels with the least amount of coverage. Pedagogical recommendations and implications of findings, including direct instruction of high frequency verbs at word levels (1,000-5,000) with the largest gaps, teaching of frequent but alternative forms for nouns and verbs present in the Frequency Dictionary, and extensive reading (Day and Bamford: 2002) at the advanced level will follow a discussion of the results. Testing theoretical word families to measure the vocabulary knowledge of native and non-native Arabic speakers and the relationship between assessments of vocabulary and reading proficiency will also be discussed as directions for future research, as well as strategies to decrease the “learning burden” (Nation: 1990, cited in Young: 2012) for learners of Arabic as a Foreign Language.

This study will venture into new territory not only in applying the corpus-based Lexical Approach to evaluating a popular TAFL textbook, but also in addressing some of the unanswered essential questions posed by Nation (2001) to define vocabulary knowledge and priorities for learning. These three primary questions remained unaddressed in TAFL studies, including how many words exist in the language at large, how many words native speakers know, and how many words are necessary to utilize the language (p. 6). As there is little data
supporting the exploration of the first two questions and only tentative steps have been taken to address the first in the context of acquisition of Arabic by non-native speakers (Van Mol: 2000, 2006), this study represents a new approach to examining these larger questions. By asking what kind of lexical knowledge a non-native learner can possess after completing Part One of the most popular Arabic as a Foreign Language textbook series, it is possible to examine how close this knowledge comes to providing students with the lexical tools necessary to eventually access authentic texts in a variety of genres.
CHAPTER 1

THE IMPACT OF DIGLOSSIA, ORTHOGRAPHY, MORPHOLOGY, AND SEMANTIC TRANSPARENCY ON THE ACQUISITION OF VOCABULARY AND READING ABILITIES IN MODERN STANDARD ARABIC (MSA)

Arabic is a member of the Central branch of the Semitic language family. As a result, it contains the corresponding structural features of this language group. A root-and-pattern based morphology and a large consonantal system consisting of guttural and emphatic consonants and a limited vocalic inventory are the primary signifiers of Arabic’s membership in this language group (Watson: 2002, pp. 1-6). It is considered to be the native language of approximately 250 million speakers and is the official language of twenty countries as well as one of the official languages of the United Nations (Holes: 2004, p. 1). Multiple varieties of the language exist, and despite structural linguistic differences, are all often considered “Arabic” by native speakers. The Classical variety of the language is the language of the Quran, serving as a liturgical language for Muslims worldwide and is viewed as the “linguistic jewel in the Islamic cultural patrimony (Holes: 2004, p. 5). Classical Arabic is lexically differentiated from its modern counterpart, commonly referred to as Modern Standard Arabic (MSA); however, both varieties are referred to by native speakers as fuṣḥa, or the “eloquent language” (Holes: 2004, p. 5). MSA is the variety most commonly taught in American institutions. According to estimates from 2009, this variety has approximately 35,000 American learners at the post-secondary level (Modern Language Association: 2010, p. 2). The linguistic or sociolinguistic reality of Arabic use does not reflect traditional teaching practices within the American academy. The linguistic distance between spoken and formal Arabic, termed “diglossia” (Ferguson: 1959), emerges as a distinguishing feature of
the Arabic language situation and poses a variety of challenges to native and non-native acquirers of MSA.

**Diglossia**

Borrowed from the French *diglossie*, the term was first highlighted in the English sociolinguistic literature by Ferguson (1959). The relationship between the two language codes entails the employment of two distinct language varieties within the same speech community, distinguished from bilingualism and bidialectism by a number of factors. In a seminal article, Ferguson defined the linguistic situation of Arabic as well as Swiss-German, Modern Greek, and Haitian Creole as:

“…a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), 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” (Ferguson 1972 [1959]: p. 345).

According to this framework, the two primary varieties of Arabic employed by native speakers are *fuṣḥa*, “the eloquent” language, used to describe Classical and Modern Standard varieties (often perceived as High) and *ṣamiyya*, the “common” language, or spoken/colloquial Arabic (often perceived as Low). Ferguson presents these varieties as diametrically opposed based on function, patterns of acquisition, and structural features. This theoretical framework describes Modern Standard Arabic (MSA), the variety long preferred by TAFL curriculum designers and professionals in American academic institutions as the language of texts, newspapers and modern written communication. This variety is appropriate for official communication, including religious sermons, political speech, news
reporting and writing, and poetry. News broadcasts are generally delivered in some form of MSA. This variety enjoys a great deal of prestige as a result of its connection to a large body of literature. In this case, the prestige of *fuṣḥa* connects with religious heritage, as the status of Arabic as the language of the Qur’an further solidifies its superior status in Ferguson’s estimation (1959: pp. 237-238). As a consequence of this prestige, the High variety is inevitably standardized through prescriptive and descriptive texts and is only acquired through formal education. As a result, it is impossible for any native speaker to acquire *fuṣḥa* as a first language (L1). The role of the mother tongue language falls to the “Low” variety, and functions as the language of everyday activity and conversation. Its employment is appropriate in everyday conversation, giving instructions to workers or servers, captions on political cartoons, and in some “folk literature” contexts. In introducing her chapter on diglossia, Bassiouney aptly characterizes the linguistic reality of diglossia in the Arabic-speaking world as one of “tension” in which “people speak one language variety at home and learn a different one in school, write in one language and express their feelings in another, memorize poetry in one language and sing songs in another” (Bassiouney: 2009, p. 9). To further complicate matters, spoken varieties of Arabic differ greatly according to degree of urbanization and geography. Spoken varieties can be specified as sedentary or Bedouin, urban or rural, and/or according to region. Regional dialects are generally divided into five primary groups: Arabian Peninsula, Mesopotamian, Syro-Lebanese or Levantine, Egyptian, and Maghrebi/North African dialects (Versteegh: 1997). An underappreciated consequence of this tension is a plurality of lexical items that fall under the umbrella of “Arabic” despite
the relative degree of divergence between these two primary varieties. Plurality of lexicon results in a challenge to native Arabic speakers in acquiring MSA as well as foreign learners.

Ferguson includes a description of linguistic features distinguishing utterances as “High” or “Low”, including phonological and morphosyntactic considerations. While both grammatical complexity and phonology are important factors in differentiating these varieties, Ferguson argues lexicon within a diglossic language situation uniquely defines both language varieties. While a large quantity of vocabulary is shared between MSA and its spoken counterparts, the employment of certain lexical items can mark an utterance as “High” or “Low”. MSA contains technical and specialized vocabulary without spoken or dialectal counterparts, whereas spoken varieties contain “popular expressions” and vocabulary for “very homey items” without corresponding terms in MSA (Ferguson: 1959, p. 242). While a similar divide between formal and informal language exists in some form within non-diglossic language communities, one of the essential characteristics of Arabic and its diglossic kin is the existence of defined “lexical doublets.” These paired items generally reference the same meaning and are in common usage; however, usage of either item brands the speech as “High” or “Low.” For example:

<table>
<thead>
<tr>
<th>MSA</th>
<th>Lebanese Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>3PMS Imp. “to speak”</td>
<td>takalam</td>
</tr>
<tr>
<td>Coordinating conj “but”</td>
<td>faqat</td>
</tr>
<tr>
<td>Adv of time “yesterday”</td>
<td>?ams</td>
</tr>
<tr>
<td>Derived Noun “breakfast”</td>
<td>fuṭu:r</td>
</tr>
</tbody>
</table>

Ferguson posits this sociolinguistic situation is mostly stable, with “relatively uncodified, unstable, intermediate forms” utilized to negotiate between High and Low varieties (Ferguson: 1959, p. 233). Later models attempted to include these intermediate
varieties. Blanc (1960) alternatively defines the reality of spoken forms according to five levels: *plain colloquial, koineized colloquial, semi-literary/elevated colloquial, modified classical, and standard classical* (Blanc: 1960, p. 110). In a study of language variation among educated speakers of Palestinian, Syrian, and Iraqi extraction, these levels are differentiated by specific phonological and morphological features, but changes from “plain colloquial” and between other levels are most numerous in terms of lexicon. Borrowings from Classical and other spoken varieties often replace spoken equivalents and concepts with no known spoken equivalent (Blanc: 1964, p. 110). Badawi’s (1973) well-known framework similarly reinterprets the bipolar sociolinguistic framework proposed by Ferguson to account for the existence of these intermediate forms. Five main varieties are posited as those most likely to be encountered by a native speaker in Egypt: *fuṣḥa a-turāth* (heritage classical), *fuṣḥa al-ṣa:ṣr* (contemporary classical), *ṣa:miyyat al-muθaqqaqāfi:n* (“colloquial of the cultured”), *ṣa:miyyat al-mutanawwāri:n* (colloquial of the basically educated), and *ṣa:miyyat al-ʔumiyyi:n* (colloquial of the illiterate) (Badawi: 1973, cited in Bassiouny: 2009, pp. 14-15). The sum of linguistic variation present within the same speech community creates challenges to both native speakers and non-native speakers when faced with the task of acquiring Modern Standard Arabic.

**Impact of Diglossia on MSA acquisition by native Arabic speakers and non-native learners**

In describing challenges native Arabic speakers face in acquiring MSA, Ibrahim (1983) characterizes the distance between MSA and spoken varieties as comparable to the gap between Latin and modern French or Italian (p. 509). Recent developments in research techniques support this widely acknowledged claim. Semantic priming tests conducted by
Ibrahim (2009) with native Arabic speakers in Israel showed similar processing effects of Hebrew and MSA, concluding that MSA is stored in the mental lexicon like a second or foreign language. Research into the practical aspects of children first encountering the MSA lexicon points to similar conclusions regarding linguistic distance. A qualitative study by Dakwar (2005) of first, second, and third grade Palestinian Arabic speakers reported translating lexical items from MSA to Palestinian Arabic as an important strategy during reading exercises (p. 81). Effects of diglossia on vocabulary acquisition point to lack of access to grade-appropriate vocabulary in Modern Standard Arabic, with effects worsening with age (Fedda and Oweini: 2012). While non-native speakers acquiring MSA as a foreign language do not face these same challenges initially, Arabic’s diglossic nature presents an array of issues to Arabic learners and curriculum designers. The very topic of which spoken variety should be taught and which combination or exclusion of certain varieties remains a point of contention among TAFL specialists.

Al-Batal (1992) details the myriad of historical and contemporary approaches to teaching Arabic: the classical approach (focused on reading Classical texts), the MSA approach (focused on utilizing MSA as the language of instruction), the Colloquial approach (where the language of instruction is spoken Arabic), the “Middle Language” approach (where the intermediate variety between MSA and the dialects is the language of instruction,) advocated by Ryding (1991) and Wilmsen (2006), and the Simultaneous approach (where both varieties are introduced at the same time). Adoption of the Simultaneous approach may create “confusion” for beginning learners that resembles the confusion native speakers face in acquiring MSA; however, this feeling could be considered “part of the whole experience
of learning Arabic” (Al-Batal: 1992, p. 302). Despite research presenting evidence in favor of an increased focus on integrating spoken varieties within American AFL classrooms (Ryding: 1991; Wilmsen: 2006; Palmer 2007, 2009), it remains an “open secret” in the TAFL profession that the language of instruction is not the language of everyday conversation (Wilmsen: 2006, p. 125). Aside from the essential challenges of mitigating the manufactured language provided by MSA-only instruction, language-specific orthography and morphology pose further obstacles to achieving automatic word recognition and efficient visual processing in Arabic as a Foreign Language.

Alphabet and Orthography: Challenges to Processing

The Arabic alphabet consists of 28 letters and is read from right to left. Short vowels (a, i, o) are not considered part of the alphabet but make up three of the six vowels in Arabic, as the other three vowels are lengthened versions. Letters exist in four positions: initial, medial, final, and independent. Reproducing the alphabet may provide little challenge in itself, but greater challenges emerge when words are transformed from characters to input. As words include only long vowels and consonants, short vowels are not part of the Arabic alphabet and are not indicated within words or as morphosyntactic markers except in religious texts and children’s books. Moreover, the popular textbook series for English-speaking Arabic learners, Al-Kitaab fii Ta’allum Al-’Arabiyya (Brustad, et. al: 2004) avoids the use of short vowels except in vocabulary lists at the beginning of each chapter. This absence of short vowels is known as “deep orthography”, in which short vowels are not evidenced at the surface but are accessible by readers with appropriate background knowledge of MSA’s features (Abu-Rabia: 2002, p. 300). In texts with deep orthography,
short vowels are also absent in the word-final position. These short vowels in word final position serve the morphosyntactic function of case marking and indicate function of words as nominative, genitive, or accusative. The absence of these vowels requires native as well as non-native readers to “deduce them through relying on context and/or linguistic prior knowledge: grammar (morphology and syntax) and early exposure to print.” Thus, poor readers may have a more difficult time comprehending Arabic texts, as many words are homographic, possessing the same written form but with a different meaning (Abu-Rabia: 2002, p. 300). A study by Hansen (2010) examining orthographic processing by non-native learners of MSA using tests of pseudowords, reading of vowelled texts, and knowledge of morphological patterns revealed the Arabic script itself slows processing speeds. As only consonants and affixes are shown in normal texts, unskilled readers may be unable to distinguish content lexemes without adequate knowledge of Arabic morphology or syntax to make sense of the surrounding context. The decoding process involved in reading MSA gives even greater importance to mastery of MSA’s morphological rules for foreign learners.

*Root and pattern: the morphological base of MSA*

Aside from orthography, Arabic is distinguished from the first languages of many American AFL learners by the Semitic root-and-pattern system. Through this system, three or sometimes four consonants are given meaning through applied consonant and vowel patterns, the most basic of which is C1C2C3 (Holes: 2004, p. 99). Generally, words containing the same root form “lexical sets” and possess some semantic relation to the root’s basic quality or core meaning. For example, words containing the root *k-t-b* possess some association with the concept of “writing”-thus, *kita:b* (book), *ka:tib* (writer), *maktaba*
(library, literally “a place for books”), and *muka:taba* (subscription) contain related meaning (Holes: 20004, p.99-100). However, deriving semantic meaning from the root and pattern system is not an exact science. Learners may experience more difficulty due to the irregular nature of semantic meaning assignment to words of the same root. Thus, *jama:l* (beauty), *jamal* (camel), and *jumla* (sentence) have little semantic association despite a shared root \[j-m-l\]. Secondly, the existence of “semivowels” /w/ or /y/ within roots may prove difficult to foreign learners, especially in common verbs such as *qa:l* (to say) and *ka:n* (to be). They must memorize at least the active present stem or verbal noun to determine how the vowel transforms within a variety of grammatical contexts. Knowledge of Arabic derivational morphology allows foreign learners to advance their ability to decode texts as well as produce lexicon productively.

*Derivational Morphology*

Arabic derivational morphology is based on the aforementioned “root and pattern” system, by which the basic root of C1C2C3 is given meaning through insertion into a number of paradigms. While fifteen patterns technically exist, only nine patterns are used commonly aside from the “unaugmented verb root” pattern of C1C2C3. Most roots take less than seven or eight of the total patterns (Holes: 2004, p. 100). According to Ryding (2005), Arabic derivational or “lexical” morphology, usually concerned with the “principles governing word formation” differs greatly from English. The line between derivational and inflectional morphological processes meant to specify “gender, number, case, and tense” is not as clear in Arabic as English (p. 44-45). Processes of derivational morphology are always considered first, and the system of derivational morphology is discontinuous (p.47). The table below
displays the ten most common verb patterns, which are those first presented in Chapter 8 of Book One of the Al-Kitaab textbook series and explained more fully in Chapter 16.

Figure 1.2-Chart of Arabic Verb Patterns (Brustad, et. al: 2004, p. 201)

<table>
<thead>
<tr>
<th>أوراقة الفعل</th>
<th>المصدر</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(varies)</td>
</tr>
<tr>
<td>المضارع</td>
<td>الماضي</td>
</tr>
<tr>
<td>نُعِلَ</td>
<td>فُعِلَ</td>
</tr>
<tr>
<td>تُعِلَ</td>
<td>فُعِلَ</td>
</tr>
<tr>
<td>مَمَعَلَة</td>
<td>فَعَلَ</td>
</tr>
<tr>
<td>إِفَعَال</td>
<td>فَعَلَ</td>
</tr>
<tr>
<td>تُعِلَ</td>
<td>فُعِلَ</td>
</tr>
<tr>
<td>تُفَعَال</td>
<td>تَفَعَال</td>
</tr>
<tr>
<td>إِفَعَال</td>
<td>تَفَعَال</td>
</tr>
<tr>
<td>إِفَعَال</td>
<td>تَفَعَال</td>
</tr>
<tr>
<td>إِفَعَال</td>
<td>تَفَعَال</td>
</tr>
</tbody>
</table>

The basic root C1C2C3 becomes Pattern I through insertion of short vowels between C1 and C2 and C2 and C3. The combination of /a-a/ is the most frequent internal vowel pattern for Form I, indicating action (transitive or intransitive) performed by an agent, including verbs like xaraj (to go out) and qaṭaf (to cut). The /a-i/ internal vowel paradigm is less frequent, indicating an “agent that is affected in some way by its own action” as well as “mental verbs.” Thus, xasir (to lose) and samiš both take /a-i/ internal vowels (Holes: 2004, p. 101). Lastly, the /a-u/ internal vowel paradigm is even less frequent, indicating an
intransitive verb acquiring a permanent quality, as in ḥasun (to become good) or saġur (to become small) (Holes: 2004, p. 101). Measure II (C1C2C2C3) adds an intensive, extensive, or causative meaning to the root by doubling the middle root consonant. Thus, Form I kasar (to break) becomes kassar (to smash) and ṣalām (to know) becomes ṣallām (to teach, make someone know). This root indicates a causative meaning for most verbs, implying another agent acted upon the initial root pattern (Holes: 2004, p. 101). Measure III (C1v:C2C3) is exertive or “conative”, implying effort on behalf of the agent to achieve the meaning of Measure I as well as interaction with a participant (Holes: 2004, p. 102). Thus, Measure I qabal (to accept, receive) becomes qa:bal (to meet, interview). Measure IV (ʔC1C2C3) is achieved by adding a glottal stop prefix to the initial root form and is generally causative. It may sometimes overlapping in meaning with Measure II (as in xabbar and ūxbar (to apprise or inform) but sometimes with a slightly shifted meaning (ʕallām (to teach, make someone know) becomes ʕilm (to inform) (Holes: 2004, p. 102). Measure V (tC1C2C2C3) is created through adding the prefix [t-] to the root and lengthening the middle radical. It generally contains the resultative meaning of Measure II. Thus, the unaugmented third-person masculine singular past perfect ḥasun (to be good) in Measure I shifts to ḥassān (“to beautify, to improve”) to taḥassan (to become better) in Measure V (Holes: 2004, p. 103). Measure VI (tC1v:C2C3) implies a plural subject and meaning is generally extended from Measure III but shows a reciprocal relationship between the subjects, as in the case of ra:sal (to correspond) and tara:sal (to correspond with each other). In other cases, Measure VI can also express stative (as in masak (to hold) and tama:sak (to hold together, cohere) or “simulative” meaning (as in jahil (to be ignorant) and tja:hal (to pretend to be ignorant)
Measure VII (nC1C2C3) possesses a pseudo-passive meaning of Measure I, as no agent is indicated (for example, *qatās* (to cut) becomes *inqatās* (to be cut off) (Holes: 2004, p. 104). Measure VIII (C1tC2C3) is derived through inserting the infix [t] between C1 and C2, and generally displays a reflexive meaning of Measure I, for example *hamam* (to worry, concern oneself) becomes *ihtmam* (to concern oneself about, to be important to s.o.) (Holes: 2004, p. 104). Measure IX (C1C2C3C3) is generally uncommon and is limited only to colors and physical attributes (as in *ḥamar* (to be red) and Measure IX *iḥmarr* (to turn red). Measure X (stC1CsC3) is much more common and productive, and is formed through adding the st- prefix to the basic root pattern of Measure I. The meaning is generally benefactive (Measure I *xaraj* (to go out) extends to *istaxraj* (to extract). Measure X can also be estimative (*baṣad* (to be far) becomes *istabṣad* (to regard as unlikely, rule out) (Holes: 2004, p. 105). Other verbs patterns (X-XV) are rare in MSA and mostly intensive.

Outside of the triliteral root system, quadrilateral verbs containing four root letters on the patterns of C1C2C3C4 (as in *tarjam* (to translate), C1C2C1C2 (like *zalzal* (to shake), C1C2C1C3 (*qahqar*, to retreat), and C1C2C3C3 (*zağlal*-to dazzle). These forms are a common and productive means for coining words for new concepts and borrowing words into the Arabic lexicon (Holes: 2004, p. 105), such as *ʔaksad* (to rust, literally ‘become oxidized’). The distinguishingly formulaic nature of derivational verb morphology similarly exists in morphological processes for noun derivation.

Derivation of nouns in Arabic is mostly systematic and thus is similar to derivation of verbs. For the equivalent of the verbal noun or gerund in English, there are a number of predictable patterns for each verb form. With the exception of Pattern I, which contains over
44 possible patterns for verbal nouns, generalizations can still be made for some forms of Pattern I and Patterns II-X. For Form I, common patterns such as CuCu:C (intransitive verbs of motion, like *xuru:j*, or “going out”), CiCa:Ca (professions and crafts, like *tija:ra* or trade), CuCu:Ca (abstract or concrete qualities, like *suhu:la*, or easiness), and CuCa:C (illness or disability, like *suda:ʔ* or headache) (Holes: 2004, p. 146-147). Patterns II through X and the quadrilateral verbs display similar predictability. The patterns of verbal nouns for these forms are listed below:

**Figure 1.3-Verbal nouns for triliteral and quadrilateral roots**

II: C1C2C2C2C3 (*fallam*, teach, make someone learn) + [taCCi:C] → tašli:m (teaching)

III. C1v:C2C3 (*qa:bl*, to meet or interview s.o.) + [muCa:CaCa] → muqa:bila (interview/ing)

Alternative Form III Pattern: (*na:qaš*, to discuss s.th. w/ s.o.) + [CiCa:C] → nīqa:š (discussion/discussing)

IV: ?C1C2C3 (*palam*, to inform) + [ʔiCCa:C] → ʔila:m (informing)

V: tC1C2C3 (*taqaddam*, to progress) + [taCaCCuC] → taqaddum (progress)

VI: tC1v:C2C3 (*tara:sal*, to correspond w/ each other) + [taCa:C] → tara:sul (corresponding)

VII: nC1C2C3 (*inqalab*, to overthrow) + [nCiCa:C] → ingila:b (overthrowing/overthrow)

VIII: C1tC2C3 (*ihtmm*, to be important to s.o.) + [CtiCa:C] → ihtmima:m (importance)

IX: C1C2C3C3 (*ihmarr*, to become red) + [CCiCa:C] → ihimira:r (reddening, blushing)

X: stC1CsC3 (*istabaʃd*, to rule out, regard as unlikely) + [stiCCa:C] → istibša:d (ruling out)

**Quadriliteral roots**

C1C2C3C4 (*trjm*, translate) + [CaCCaCa] → tarjama (translating, translation)

C1C2C1C2 (*zklz*, shake) + [CiCCa:C] → zilza:l (earthquake)

Standard Arabic also possesses a systematic method for deriving active and passive participles. The active participle is the entity that “is, was, and will be” performing an action, has performed the action, or habitually performs the action. Although the active participle is

---

1 Verbs with a semivowel /w/ or /y/ as a root letter in Forms IV and X take an –a suffix; “weak verbs” containing a semivowel as the third radical also take –a suffix.

2 Adapted from Holes: 2004 (p. 147)
a noun, it can also function as a verb or adjective. The passive participle instead expressed the entity that results from the action and functions as a “quasiverbal adjective” (Holes: 2004, pp. 149-150). Each verbal pattern possesses a systematic way to derive participles. Only Form I possesses a distinct pattern, in which the unaugmented root (for example, *ktb*, write) is inserted into the pattern Ca:CiC for the active participle and maCCu:C for the past participle. Thus, *ka:tib* (writer) and *maktu:b* (written) are derived in this manner. Patterns II-X and quadrilateral patterns are more closely related, as the derivation of participles for these forms involves the simple process of subtracting the prefix /yu-/ from the 3rd person masculine singular p-stem verb and affixing /mu-/ to the form. The “theme vowel” is then modified to /i/ for active participles and /a/ for passive participles.

Figure 1.4-Deriving active and passive participles

Form II (*ʔllm*, to teach)

<table>
<thead>
<tr>
<th>Active Pattern</th>
<th>Passive Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>yuʔallim</em> (3PSG p-stem V) /yu p-stem + AP template [muCaCCiC]</td>
<td><em>muʔallim</em> (teacher)</td>
</tr>
<tr>
<td><em>yuʔallim</em> (3PSG p-stem V) /yu p-stem + PP template [muCaCCaC]</td>
<td><em>muʔallam</em> (s.th. learned)</td>
</tr>
</tbody>
</table>

“Primitive nouns” are not verbally derived and can therefore only function as nouns. Lexemes such as *rajul* (man) and *ma:ʔ* (water) are examples of this word type. Nouns of this type may cause problems for foreign learners, as gender is not always apparent. While the gender of words like *luğa* (language) and *dašwa* (claim) can be recognized as feminine due to the –a suffix, learners must memorize the gender of words like *harb* (war) and *šamš* (sun), which are feminine by convention (Holes: 2004, p. 155). Primitive nouns are differentiated from derivatives, which are nouns formed from other nouns and verbs. This type of noun may cause less trouble for learners once the verb measures are mastered, as the formation of this category involves manipulation of the basic derivational forms for nominalization.
Figure 1.5- Derived nouns

*Noun of instance* (verbal noun + -a suffix)

Qafž (jumping) + -a suffix → qafža (a jump)

*Noun of manner* (CiCCa)

Jalas (sit) + [CiCCa] → jilsa (a manner of sitting)

*Noun of place* (maCCaC, MaCCiC, or maCCaCa)

ṣanaṯ (manufacture) + [maCCaCa] → maṣnaša (factory, lit. “place for manufacturing”)

*Noun of instrument* (miCCaC, mica:C, or MiCCaCa) and occupation (CaCCa:C, CaCCa:C)

fataḥ (open) + [mica:C] → miftaḥ (key, lit. “instrument for opening”)

kaḍab (lie) + [CaCCa:C] → kaḍda:b (liar, someone who lies continuously)

*Diminutives* (CuCayC(a), CuCayyic(a))

kalb (dog) + [CuCayC] → kulayb (small dog)

*Inflectional Morphology*

The learnable and formulaic nature of the basic forms within derivational verbal morphology may be also considered to transfer to inflection for person, gender, number, and mood. The aforementioned derivational forms provide the basic pattern of perfect verbs, which are transformed into imperfect, subjunctive, and jussive verbs through formulaic processes of affixation and insertion of short vowels. Only the perfect, imperfect, subjunctive, and jussive are considered verb patterns in Standard Arabic, with the subjunctive, jussive, and imperfect gathered together under the umbrella of imperfect verbs. Briefly, this section will examine passive and active imperfect/perfect verbs to build a case for the learnable nature of Standard Arabic verb morphology.

In examining the relationship between perfect and imperfect verbs, Holes (2004) considers perfect verbs to fall under the category of “suffix stem”, through which verbs are

---

3 Examples and templates adapted from Holes: 2004, pp. 155-160
inflected for person, mood, gender, and number through appending a suffix to the end of the derived root in question. In contrast, he defines present tense or “imperfect” verbs as “p-stem”, through which information about the verb is inflected through adding a prefix. The p- and v-stem versions of Pattern I (C1C2C3) contain predicable internal vowel patterns for the active and passive voice. The first vowel is always /a/ in the active voice and /u/ in the passive. While s-stem verbs with a second vowel as /a/ may manifest as /a/, /i/, or /u/ when transformed into p-stem verbs, a few rules govern these changes. If C2 or C3 is guttural, the second p-stem vowel is usually /a/. The occurrence of /i/ in position 2 (v2) in s-stem generally becomes /a/ in the p-stem, and /u/ in this position usually corresponds to /u/ in the p-stem/ (Holes: 2004, p. 107). Other verb measures similarly display the same predictability, with basic paradigms for internal vowelling shown when patterns are grouped in a systematic way. In the active voice, Forms VII, VIII, IX, and X contain a /a-a-i/ vowel pattern for the p-stem and /a-a/ pattern for the s-stem. The passive rule corresponds to /u-a-a/ for p-stem verbs and /u-i/ for s-stem verbs. Forms V and VI display a simple paradigm of /a-a-a-a/ for active p-stem verbs and /a-a-a/ for active s-stem verbs. Passive verbs in these patterns become /u-a-a-a-a/ in the p-stem and /u-u-i/ for s-stem verbs. Forms II and III contain the /u-a-i/ vowel pattern for active p-stem verbs and /a-a/ for active s-stem verbs. Passive verbs in these patterns transform into /u-a-a/ in p-stem and /u-i/ for s-stem. Lastly, form IV displays a pattern of /u-i/ for active p-stem verbs and /a-a/ for active s-stem verbs. Verbs are made passive through altering vowels to /u-a/ in the p-stem and /u-i/ in the s-stem (Holes: 2004, p. 108). These patterns are then altered for person, gender, and number. Despite the expanded number of pronouns, including a second and third-person masculine and feminine pronoun
that exists in the singular, plural, and dual forms, the formulaic nature of verbal derivation and inflection allow for a system that can be learned with consistent effort and practice on the part of foreign language learners. Noun morphology displays a similar pattern of predictability despite some complexity.

*Adjectives, number, agreement, and case*

In MSA, adjectives may be derived relationally (attaching /i:/ to nouns) or through use of comparative and superlative on the pattern ?aCCaC for elatives (as in ?akbar, the biggest of greatest, or bigger or greater) and the feminine CuCCa: (as in kubra:, greater or bigger) (Holes: 2004, p. 159-160). Adjectives must agree in gender and number with the nouns they modify, including definiteness (marked by the al- prefix), as part of a construct phrase, by affixation of pronoun, or as a proper name. Feminine nouns ending in –a (ta? marbu:ta) are pluralized by the suffix –at. This aspect of Arabic may prove difficult for beginners, as English does not require this sort of concord between nouns and adjectives.

Noun plurals are formed either through as “suffixization” (for sound plurals) or “interdigitation” (broken plurals) (Holes: 2004, p. 162). The system for deriving broken plurals may seem challenging for learners with little background knowledge; however, these patterns can be learned and encounters with previously unknown broken plurals should become easier with the passage of time. The case marking system on the other hand may not prove as easy to acquire without some effort. Known as i’ra:b, it is a morphosyntactic system for specifying case of words as nominative, accusative, and genitive. Definite nouns are marked with short vowels /u/ (nominative), /a/ (accusative), and /i/ (genitive) and indefinite nouns are marked by their corresponding definite counterparts /-un/, /-an/, and /-
The suffixes of sound plurals must be specified according to these parameters in both dual and plural forms. Students often complain this system is difficult to learn; however, the systematic nature of this feature allows its acquisition with some effort. Other features of Arabic, such as syntax and phonology, may pose different difficulties for new as well as advanced learners.

Other challenges: “degree of transparency” and degree of synonymy

Lastly, acquisition of MSA vocabulary is hampered by the slight degree of “linguistic transparency” between English and Arabic (Van Mol: 2006, p. 306). The lack of cognates and intransparency of borrowed or foreign words increases the sheer size of vocabulary needed to be acquired by foreign learners, relative to other commonly taught Latin-based languages. Approximately one in 1,000 words is a cognate in both Arabic and English, and borrowed words such as demokratyya (democracy) may not be automatically recognized by foreign learners (Van Mol: 2006, p. 306). On the subject of coined and borrowed words, Baker (1987) notes the gap between suggestions made by Arabic language academies, varied methods for coining or adopting new Arabic lexicon, actual language use, and the multitude of Arabic varieties creates a plethora of synonyms. The sum of these challenges creates more than a few obstacles to acquiring proficiency in MSA as a foreign language. To provide a clearer picture of how hypotheses concerning meaning, organization of the mental lexicon, and vocabulary knowledge and reading interact, this paper will next examine the literature concerning fundamental questions regarding word knowledge in the SLA and psycholinguistic literature in general as well as in the specific TAFL context.

---

4 As an exception to the rule, diptotes do not take /un/, /in/ or /an/ when indefinite, with /a/ sufficing to mark both genitive and accusative cases.
CHAPTER 2
DEFINING WORDS, THE MENTAL LEXICON, AND THE LINK BETWEEN READING AND VOCABULARY

What is a word: Units of counting and organization of the mental lexicon

According to Nation and Meara (2002), defining the basic linguistic unit of a “word” is one of the greatest challenges to vocabulary research. Depending on the goals of researchers, tokens or types can be counted. In counting tokens, every discrete word is counted. A sentence like ʔakala alwaladu a-ittufāha (“The boy ate the apple”) contains three tokens or five if the prefix al- (definite article “the”) is counted even as same sentence contains five tokens in English. An alternative unit of counting is the lexeme or lexical item, or “an item that functions as a single meaning unit, regardless of the number of words it contains” (Schmitt: 2000, p. 2). Specifically addressing items containing multiple tokens, “multi-word units” are considered to be “learned, stored, and used as complete units” (Nation and Meara: 2002, p. 36). Therefore, phrases like ahlan wa sahlan (roughly, “welcome”; “be of our family and be at ease”, or “make yourself at home”) and ṣabāḥ al-xeir (“good morning”; literal translation: “morning of goodness”) are treated as one unit regardless of the number of tokens the unit contains. Alternatively, the lemma, which includes only the base word and all of its inflected forms, may be used as a “conservative” unit of word knowledge (Schmitt: 2000, p. 2). In Arabic, sound masculine and feminine plurals derived from ṭa:lib (masculine singular active participle/adjective) such as ṭa:liba (feminine singular active participle/adjective), ta:libat (feminine plural active participle/adjective), and the “broken” or discontinuous plurals like ṭula:b (discontinuous masculine plural of the original active
participle) or talaba (an alternate discontinuous plural of the original active participle) belong to the same lemma despite morphological alterations.

However, if the unit of counting is supposed reflect “the kind of knowledge that language users draw upon”, then the most productive unit of counting is the word family (p. 36). Bauer and Nation (1993) define word families in English as a “base word and all of its derived forms that can be understood without having to learn each form separately”, where knowledge of the base and derived forms allows for recognition of other family members “with little to no extra effort” (p. 253). The meaning of the base word and the derived or inflected form must be close in order to be members of the same family. Thus, “walk”, “walker”, and “walking” are members of the same family while “hard” and “hardly” are not (Ibid, p. 253). Eight criteria are suggested for classification of levels within words families: frequency of affixes, productivity, predictability, regularity of the written base form, regularity of the written base, regularity in spelling of the affix, regularity in the spoken form of the affix, and regularity of function of the affix (Ibid, p. 256). Seven levels of affixes indicate development of a learner’s lexicon in English. At level one, the learner is unable to recognize the difference between “book” and “books”, but by level six the learner should recognize “frequent but irregular affixes” such as –able, -ee, -ic,-ify, -ion, and –ist (Bauer and Nation: 1993, p. 261).
Figure 2.1- Additions to a word family at different levels of inflection and affixation (Bauer and Nation: 1993, p. 254).

<table>
<thead>
<tr>
<th>Levels</th>
<th>Word families</th>
<th>Wood</th>
<th>Wood's</th>
<th>Woods</th>
<th>Wooden</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>develop</td>
<td>wood</td>
<td>wood's</td>
<td>woods</td>
<td>wooden</td>
</tr>
<tr>
<td></td>
<td>develops</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>developed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>developing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>developable</td>
<td>woody</td>
<td>woodiest</td>
<td>woodier</td>
<td>woodiness</td>
</tr>
<tr>
<td></td>
<td>undevelopable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>developer(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>undeveloped</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>development(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>developmental</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>developmentally</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>developmentwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>semideveloped</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>antidevelopment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>redevelop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>predevelopment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>anti-wood</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

While these seven levels may allow for ease of classification in English, the morphological system of derivation in Arabic may diminish the need to rely on classification of word families according to these levels. The development of word families for foreign learners of Arabic may be instead hypothesized according to acquisition and control over morphological patterns as well as semantic relationship between the initial root and derived word. An examination of the root *k-t-b* in *A Dictionary of Modern Written Arabic* (Cowan and Wehr: 1994), colloquially referred to as “Hans Wehr” after the dictionary’s initial author, reveals very few derived tokens that would be considered members of separate families.
### Figure 2.2 - Word families for *k-t-b*

**Family One - k-t-b**

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Transliteration</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>write (3PMSG, Perfect)</td>
<td>kataba</td>
<td>كتبَة</td>
</tr>
<tr>
<td>to make s.o. write</td>
<td>kattaba</td>
<td>كتبٌ</td>
</tr>
<tr>
<td>to correspond (with s.o.)</td>
<td>kataba</td>
<td>كتبٌ</td>
</tr>
<tr>
<td>to dictate/make s.o. write</td>
<td>aktaba</td>
<td>أكتب</td>
</tr>
<tr>
<td>write letters to each other</td>
<td>taka:taba</td>
<td>تكتاب</td>
</tr>
<tr>
<td>to subscribe</td>
<td>inkataba</td>
<td>أكتب</td>
</tr>
<tr>
<td>to copy (make s.th. written)</td>
<td>iktabata</td>
<td>أكتب</td>
</tr>
<tr>
<td>to ask s.o. to write; dictate</td>
<td>istaktaba</td>
<td>استكتب</td>
</tr>
<tr>
<td>Quranic school</td>
<td>kutaab</td>
<td>كتبَة</td>
</tr>
<tr>
<td>books</td>
<td>kulub</td>
<td>كتبَة</td>
</tr>
<tr>
<td>book seller</td>
<td>kita:bi</td>
<td>كتبَة</td>
</tr>
<tr>
<td>writing</td>
<td>kita:bi</td>
<td>كتبَة</td>
</tr>
<tr>
<td>written</td>
<td>kita:bi</td>
<td>كتبَة</td>
</tr>
<tr>
<td>office</td>
<td>maktab</td>
<td>مكتبة</td>
</tr>
<tr>
<td>office-related</td>
<td>maktabi</td>
<td>مكتبة</td>
</tr>
<tr>
<td>library</td>
<td>maktaba</td>
<td>مكتبة</td>
</tr>
<tr>
<td>typewriter</td>
<td>mikta:bi</td>
<td>مكتَّب</td>
</tr>
<tr>
<td>correspondence</td>
<td>muka taba</td>
<td>مكتَّب</td>
</tr>
<tr>
<td>subscription</td>
<td>iktita:bi</td>
<td>أكتب</td>
</tr>
<tr>
<td>dictation</td>
<td>istiktab</td>
<td>استكتب</td>
</tr>
<tr>
<td>dictaphone</td>
<td>istiktabi</td>
<td>استكتب</td>
</tr>
<tr>
<td>writer (masc. and fem.)</td>
<td>kati:bi, kati:ba</td>
<td>كاتِب, كاتِبة</td>
</tr>
<tr>
<td>written; destined; failed</td>
<td>maktu:b</td>
<td>مكتوب</td>
</tr>
<tr>
<td>writing; messages note</td>
<td>maka ti:b (plural)</td>
<td>مكتِّب</td>
</tr>
<tr>
<td>correspondent</td>
<td>muka tib</td>
<td>مكتَّب</td>
</tr>
<tr>
<td>subscriber</td>
<td>muktatib</td>
<td>مكتَّب</td>
</tr>
<tr>
<td>booklet</td>
<td>kutayyib</td>
<td>كتبَة</td>
</tr>
</tbody>
</table>

**Family Two - k-t-b**

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Transliteration</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>library/bookstore</td>
<td>kita:bk:na, kutubk:na</td>
<td>كتابخانَة, كتبخانَة</td>
</tr>
<tr>
<td>(borrowed from Persian)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Family Three - k-t-b**

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Transliteration</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>squadron/battalion/corps</td>
<td>kati:ba</td>
<td>كتبَية</td>
</tr>
<tr>
<td>squadrons/battalions/corps</td>
<td>kata:bi</td>
<td>كتبَة</td>
</tr>
<tr>
<td>pertaining to the Lebanese Phalange Party</td>
<td>kata:bi</td>
<td>كتبَة</td>
</tr>
</tbody>
</table>
Importantly, these word groupings are thought to be “psychologically real” units (Nagy, et.al: 1989; Bertram, Laine & Virkkala, 2000, cited in Nation: 2012). The following section will explore general theories for differences between the L1 and L2/FL “mental lexicon” (VanPatten and Benati: 2010) as well as insights from the context of acquisition of Hebrew and Palestinian Arabic and general mental lexicon studies to present a hypothesis for the basis of Arabic word families.

Defining the mental lexicon

Lexicon, or the sum of words known in a language by a given individual, operates like a “mental dictionary” of words, containing “their meanings, the sounds that make up words, and required and optional semantic and grammatical information” with links to other words and concepts (VanPatten and Benati: 2010, p. 103-104). While apt for its poetic description of the size and content of a given person’s lexical knowledge, this metaphor needs to be applied with some caution. Unlike a dictionary, words found in the mental lexicon are not organized in such a linear fashion. The following section will examine theories for organization of the myriad of meanings, concepts, and words in the mental lexicon for learners of Arabic as a Foreign Language.

To date, there are no known studies on the role of the root or word/stem in the lexicon of learners of Arabic as a Foreign Language. There is a similar dearth of information regarding development of depth and breadth of lexicon throughout the learning process. Still, it is possible to offer hypotheses regarding the nature of the mental lexicon of these learners. First, this mental lexicon is distinctly different from that of a monolingual. Early attempts to describe the differences between a native L1 speaker and an L2 learner of the same language
centered on the “existence of a separate linguistic system” for L2/FL learners attempting to approximate “TL [target language] norms”, termed “interlanguage” (Selinker: 1972, p. 214). However, this model implies there is something deficient in the language system of a foreign or second language learner. The definition of “multicompetence”, as a “compound state of mind with two grammars” or a “distinct state of mind” from monolinguals provides a less pejorative basis for the model (Cook: 1992). By this account, L2 learners should be measured “at best” against the skills of a fluent bilingual, not a monolingual (Cook: 1992, p. 559). This model posits the linguistic system of an L2 learner contains a grammar, “metalinguistic awareness”, and notably, evidence of “interrelationships” between the lexicons of L1 and L2 (p. 567). Of the relatively few studies conducted to explore the development of the linguistic system of Arabic FL learners, recent studies instead examined the development of Arabic morphology in the Arabic FL linguistic system. A study of predicted and actual acquisition of verbal agreement morphology in learners of Arabic as a Foreign Language in Australia concluded lexical morphemes like past tense are acquired first, with differences between learners in speed of acquisition of morphological patterns but not order (Mansouri: 2001). Unfortunately, no known study examines or measures the lexical contents of the linguistic system of Arabic FL learners. Exploration of the role of the consonantal root and the stem or word in the Arabic FL lexicon remains uncharted territory. Still, insights from different SLA contexts may provide a basic understanding of how novice-level Arabic learners initially store and perceive items within the mental lexicon.
Stages of vocabulary acquisition and the role of the root and word

General descriptions of the L2/FL vocabulary acquisition process indicate the process may differ for novice language learners and advanced students. Jiang (2004) argues semantic development occurs in two stages for the majority of L2 words: the “comprehension” and “development” stages. While these processes are not mutually exclusive, it is generally accepted that stage one involves mapping new L2 lexical items to “pre-existing concepts or L1 translations” if there is a similar concept to the L2 target word (Jiang: 2004, p. 104). Although Jiang (2004) prefers to refer to this stage of vocabulary acquisition as the “comprehension stage”, it has also been referred to as the “acquisition stage” (Ellis: 1995, cited in Jiang: 2004, p. 104). As knowledge of the “semantic structures that are specific to L2 words”, learners begin to progress towards the “development stage.” During this stage, semantic meaning is refined and meaning of new L2 words is restructured after initial transfer from L1 (Jiang: 2004, p. 104). At this stage, the learner is thought to review the initial associations made between L1 concepts and L2 vocabulary items. In a study of reaction times for semantic judgment tests, Jiang (2002, cited in Jiang: 2004) showed ten advanced Chinese and Korean ESL students responded with faster reaction times for words with the same L1 translation and L2 English meaning, pointing to the continued mapping of L2 words to L1 concepts, even at the advanced level. However, a second test of advanced Chinese and Korean ESL learners in sentence completion tasks with correct answers lacking precise L1 translations revealed “better than chance” performance. This led the researcher to conclude accuracy level correlates to the learner’s progress in the semantic development process and semantic restructuring (Jiang: 2004, p. 118). This evidence points to the possible
reliance of novice Arabic FL learners on their L1 conceptual knowledge and a decreased likelihood of associations between words of the same root with divergent patterns as control over semantic connections between the form of Arabic’s roots and patterns and meaning may not be fully developed. Ravid’s (2005) review of studies examining the development of root awareness in native Hebrew and Palestinian Arabic speaking school-age children links successful completion of root-awareness tasks to increasing levels of literacy and formal education based on the root-and-pattern approach. The connection between the root-and-pattern basis of instruction in Arabic FL textbooks and reports of increased root awareness as a result of education may give credence to the root basis of an Arabic FL learner’s lexicon. On the other hand, the framework described by Ellis (1995, cited in Jiang: 2004, p. 104) and Jiang (2004) may indicate most Arabic learners rely on direct mapping of memorized meaning and to new Arabic forms, especially if new items cannot be analyzed morphosyntactically to allow for a more comprehensive, and therefore, restructured meaning of the concept in the L2. This supposition may lend a solider foundation to hypothesizing about the development of word families based on the consonantal root as well as the word in instance where the root is less transparent.

In light of the possible arrangement of the Arabic learner’s lexicon based on the root as well as the word, the concept of word families can be transferred but altered to meet the nature of the Arabic language, that is, the root-and-pattern basis of derivational morphology and memorized words or Multi-Word units (Nation and Meara: 2002) as presented in AFL materials. Moreover, evidence supports the importance of word families to conceptualize and track the development of the depth of students’ vocabulary knowledge. A study by Morin
(2006) in the context of Spanish as a Foreign Language revealed one semester of direct instruction in strategies to increase the size of Spanish word families such as processes of affixation, roots, and recalling parts of speech yielded significant gains for the experimental group in tests of sight vocabulary, productive knowledge, and receptive knowledge.

Knowledge of word families may also increase receptive knowledge, or knowledge necessary to perform reading and listening activities effectively within the language (Carlisle and Katz: 2006). This study will next examine the divide between the types of knowledge necessary to utilize a language.

*Dividing the mental lexicon: aspects of word knowledge and the receptive/productive divide*

Nation (1990, p. 31) defined other characteristics of word knowledge that remain integral to a fuller picture of all aspects involved in word knowledge. An outline of these concepts examined through the lexical item *qa:l* (to say) is presented below.

Figure 2.3- A sketch of word knowledge for *qa:l* (to say) (based on Nation: 1990, p. 31).

**Meaning:** to say

**Written form:**

**Spoken form:** *q* (voiceless uvular plosive) becomes *ʔ* (glottal stop) in urban Levantine and Cairene varieties; *g* (voiced velar stop) in Gulf/ruralite Levantine varieties

**Grammatical behavior:** “hollow” verb where medial vowel *a:* (low front unrounded vowel) becomes *u:* (mid-high back rounded vowel) in the present imperfect tense; underlying root letter is *w* (voiced labio-velar approximant)

**Collocations:** as a noun- *qa:l SaHi:H* (truthful speech); *qa:l wa ’amal* (word and deed)

**Register:** formal and informal; used to report speech

**Associations of the word:** *?inn* (that)

**Frequency:** high frequency in written and spoken discourse

While this framework proposes a complete coverage of word knowledge types, not all types of word knowledge are evenly distributed. A language user may possess developed knowledge of the grammatical functions, meaning, and written form of a word, but possess
little knowledge of the spoken form or associated words or register. The lack of immediate knowledge of all word knowledge aspects led Schmitt (2001) to argue vocabulary acquisition is “incremental.” “Immediate mastery” of all types of knowledge for a given lexical item is not possible (Nation: 1990, p. 5). The lack of balance implicit in word knowledge for native as well as non-native Arabic learners presents a particular challenge to classification of representation in the mental lexicon. The following section will examine a framework for distinguishing between the two primary types of vocabulary in the mental lexicon as receptive and productive in order to present a more multi-dimensional picture of word knowledge for AFL learners.

While Nation (1990) does base his model for word knowledge on the three main precepts of form, meaning, and use, he makes a further useful distinction between productive and receptive vocabulary knowledge in conceptualizing a model of second language vocabulary acquisition. According to this division, listening and reading are receptive skills involving perception of vocabulary while productive vocabulary is output-based, involving accessing specific parts of the mental lexicon for meaning-making activities (Nation: 2001, p. 25). Alternative interpretations of the divisions between vocabulary types include the “passive”, “controlled active”, and “free active” vocabulary framework (Laufer: 1998) and “passive” and “active” vocabulary division (Meara: 1990). According to Laufer’s (1998) model, “passive” vocabulary is alternatively defined as “basic receptive” vocabulary, or knowing the “most frequent or core meaning of a word.” Laufer’s framework further divides “active” (productive) vocabulary into “controlled” and “free” active vocabulary. According to this distinction, “controlled” active lexicon involves task-based production of a word while
“free” productive knowledge involves using words at “one’s free will, without any specific prompts for specific words” (Laufer: p. 257). While the latter half of the model may provide an interesting framework for conceptualizing and measuring productive vocabulary measures, Laufer uses the terms “passive” and “receptive” interchangeably; however, “receptive” is preferable to “passive” as the latter implies a less important role in vocabulary acquisition. In addressing the issue of defining vocabulary simply as “passive” and “active” vocabulary, Nation (2001) objects to this due to the associational nature of Meara’s model on similar grounds. While he concedes these terms are often used interchangeably, Meara’s model indicates passive vocabulary can be accessed only by external stimuli, or by seeing or hearing a word. On the other hand, “active” word knowledge may be called up through associations with other words or external stimuli. Nation’s (2001, p. 25) critique points to the basis of vocabulary as “meaning-driven” rather than formed by associational knowledge.

Aside from the debate concerning terminology, a recent study by Webb (2008) showed scores for complete word knowledge is larger for receptive than productive knowledge, with the size of productive and receptive knowledge generally correlating. Students with a larger receptive vocabulary are then likely to possess a larger productive vocabulary. Specific quantities of vocabulary have been hypothesized to correlate with these specific types of knowledge, as well as the ability to use the language in these knowledge specific domains. The following section will examine the quantity of knowledge necessary to access authentic materials receptively, and the role of frequency data in informing these projections.
Receptive Vocabulary: Size and importance to reading

A strong link exists between vocabulary knowledge and reading ability, as some studies (Carver: 2000, 2003, cited in Grabe: 2009, p. 266) predict an almost perfect correlation between knowledge of vocabulary and reading comprehension. In reviewing the literature on this connection, Grabe (2009) asserts there is a “strong and reliable relationship” between these two aspects of SLA (p.266). In the ESL context, Qian (1999) found especially high correlations between tests of reading comprehension and depth and size of vocabulary, with Pearson coefficient correlations for pairs of reading comprehension, vocabulary size, depth-of-vocabulary knowledge, and morphological knowledge all registering higher than 0.60 (p<0.05). While depth or quality of word knowledge plays an important role, simple breadth of word knowledge plays just as important of a role in reading comprehension.

Ability to access texts of a variety of genres is constrained by knowledge of vocabulary, especially of the most frequent words in a given language. For this reason, efforts to record and utilize frequency data are well known and well documented. Early approaches to documenting word frequency in English by Dolch (1936) created a list of 220 “sight words” that should be easily recognized by elementary school-aged children in reading and West (1953) created a “General Service List” of English vocabulary designed to assist learners of English (cited in Folse: 2004, p. 42). More recent attempts to utilize frequency information for pedagogical purposes by Xue and Nation (1984) and Coxhead (2000) resulted in the creation of the University Word List (808 frequently occurring words in academic texts) and the Academic Word List (570 word families especially designed for non-native English learners) (cited in Folse: 2004, p. 42). While many attempts to gather and
organize frequency data resulted only in long lists of words, recent lists consider the existence of word families. A retrospective evaluation of word families by Nation (2001) shows the basis of word families as considering each family as “a unit of counting” that accounts for the “kind of knowledge that language users draw on.” As previously discussed, the role of the root in organizing the mental lexicon for native Arabic speakers, and ostensibly, non-native Arabic learners provides a basis for the consideration of “word families” in Arabic, even as discrete words or lemmas have been the only unit of counting utilized in recent projections for text coverage in Arabic.

*Frequency and text coverage in English and Arabic*

Focusing on high frequency vocabulary is intricately tied to the idea of utility and language use, as “not all words are equally useful”—in English, the word “the” can account for upwards of 7% of each text (Nation and Waring: 1997). Knowledge of the even the minimal number of the most frequent word families is correlated with high rates of hypothetical text coverage. The first 1,000 most frequent word families (1,600 words) in English are thought to cover over 77% of a written text (Nation: 2001) and the same 1,000 word level in Spanish will provide 76% coverage of non-fiction texts and 79.6% of fiction texts (Davies: 2005, p. 109, cited in Godev: 2009, p. 55). Laufer (1992) recommends a minimum of 3,000 word families or approximately 5,000 distinct words for “minimal comprehension”; however, recent accounts suggest higher numbers. Hu and Nation (2000) found independent reading with the goal of learning new vocabulary without assistance from a dictionary or outside sources should maximally have 1 in 50 running words unknown to achieve comprehension, or only 2% of words being unknown. Intensive reading with a
“fairly heavy vocabulary load” equates to more than 5% of running words in the texts being unknown. The study of 66 ESL learners revealed “no one” gained “adequate comprehension” at 80% text coverage, but it was possible for some to gain “adequate or close to adequate comprehension” at 90-95%. Results indicate the “threshold” for vocabulary knowledge to affect adequate text coverage and reading comprehension lies between “80-90%”; however, the “adequate unassisted comprehension” of texts for pleasure reading may lie closer to 98%. While a more recent study (Schmitt, Jiang, and Grabe: 2011) questions the idea of a distinct threshold below which comprehension is not possible, results still suggest a gradual relationship between vocabulary coverage and reading comprehension. Results from the study of 661 ESL students revealed 95% vocabulary coverage of texts may yield approximately 60% comprehension while a more ideal rate of 70% or 100% may be linked to coverage rates of 98-99% and 100%. This logic of text coverage reveals “it does not make much sense having students read texts for which they do not know 10% or more of the words” (Schmitt, Jiang, and Grabe: 2011, p. 36). Other studies examined the actual quantity of the most frequent words or word families to reach these levels of text coverage. In investigating theoretical rates of vocabulary coverage utilizing the British National Corpus, an English corpus of 100 million tokens, Nation (2006) found that reaching the ideal text coverage rate of 98% for independent reading requires 8,000 word families to cover a newspaper text and 9,000 word families to read a novel. Lower coverage rates of approximately 95% require 4,000 word families for accessing novels and newspapers if proper nouns are also known. Relatively little is known about the relationship between
frequency and text coverage in Modern Standard Arabic. Moreover, many frequency studies are out of date.

In the introduction to the most up-to-date frequency dictionary of Arabic, Parkinson describes previous frequency studies in wide circulation as “seriously outdated” or “relatively small” (Buckwalter and Parkinson: 2011, p. 2). The earliest frequency study by Brill (1940) resulted in the “Basic Word List of the Arabic Daily Press” and was based on a corpus of only 136,000 tokens, and even later studies by Abduh (1979) and Koulou (1991) examined only 255,000 and 200,000 tokens, respectively (cited in Buckwalter and Parkinson: 2011, p. 2). A more recent effort by Oweini and Hazoury (2010) examined a corpus of 24 readers for elementary students grades K-3 to create a sight word list of 500 words as a pedagogical tool. Razak (2011) compiled a corpus of 30 newspaper articles from the World Affairs and Sports sections of seven Arabic language newspapers, amounting to over 175,000 tokens. He found a strong correlation between each newspaper section and specific sets of high frequency words specific to the subject matter of each section. The most recent dictionary complied by Buckwalter and Parkinson (2011) represents the most comprehensive collection of accessible frequency data to date in the Arabic language context. Utilizing a corpus of 30 million words, derived from 90% written materials and 10% transcribed spoken data, the dictionary indexes the 5,000 most common lemmas (words and primary inflected forms) in present usage. While recent work on frequency points to increased importance of developing measures of proficiency and text coverage utilizing natural language corpora, observations regarding proficiency levels and frequency data have been anecdotal, and only one study examined the connection between frequency, vocabulary, and text coverage.
Suppositions for vocabulary size based on “general studies on vocabulary size” and “anecdotal reports from teachers of Arabic” indicate an Advanced learner of Arabic should know 3,000-3,500 high frequency words (Al-Batal: 2006, p. 333). It is not clear whether this observation refers to a measure of productive or receptive knowledge, but other studies do address the divide between knowledge for reading or listening and speaking or writing. Data from a corpus of 4,000,000 tokens extracted from written and spoken Arabic texts suggests higher numbers.

**Figure 2.4- Conceptualizing Levels Based on Vocabulary (Van Mol: 2000)**

<table>
<thead>
<tr>
<th>Level</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level A-Preparatory Level</td>
<td>3,000 words</td>
</tr>
<tr>
<td>Level B-Oral and Written Media Level</td>
<td>8,500-10,000 words</td>
</tr>
<tr>
<td>Level C-Novels and Literature Level</td>
<td>12,000-14,000 words</td>
</tr>
</tbody>
</table>

While the above vocabulary levels refer to the link between proficiency and vocabulary knowledge in general, other projections by Van Mol (2006) refer to the specific vocabulary threshold necessary for adequate text coverage in MSA. In analyzing text coverage of Media and Literary texts, Van Mol (2006) found knowledge of the 8,500 most frequent words from a 4,000,000 token corpus covered only 95% of the most common words in Media texts. He further estimated 12,000 of the most frequent words provide 95% coverage of the vocabulary currently employed in literary texts (2006, p. 310). In comparing the results of Nation (2006) and Van Mol (2006), a large gap emerges between the lexicon needed to access authentic English and Arabic language texts. Utilizing Nation’s (1983, cited in Laufer: 1997, p. 24) formula for word family to lexical item conversion (where one word
family contains approximately 1.6 members), results from Nation’s later estimation of text coverage of newspapers at 95% coverage rate is approximately 6,400 distinct word units including proper nouns, or 4,000 word families (Nation: 2006, p. 72). The average lexicon size for reading unaltered novels (Lord Jim by Joseph Conrad, Lady Chatterley’s Lover by D.H. Lawrence, The Turn of the Screw by Henry James, The Great Gatsby by F. Scott Fitzgerald, and Tono-Bungay by H.G. Wells) is similarly estimated at 4,000 word families, or approximately 6,400 lexical items including proper nouns (Nation: 2006, p. 71). In comparing the lexical knowledge necessary to cover 95% of most newspaper texts, approximately 2,100 more distinct words or lemmas are needed in Arabic than English. In the case of novels, a much larger gap emerges. If Van Mol’s (2006) estimates are reliable, a reader attempting to access an Arabic language novel needs knowledge of 5,600 more words than when reading a novel in English, nearly double the number of discrete words necessary in English. One difficulty in interpreting this data is the lack of reciprocity between the two data sets. As the concept of word families in Arabic remains unexplored, a comparison must rely on converted data from Nation’s (2006) study. Moreover, it is not possible to verify the quantity or exact type of text utilized in Van Mol’s (2000, 2006) estimations regarding vocabulary levels and text coverage, or even what was meant by “words” in his study. Due to the relative dearth of knowledge concerning the connection between frequency data from corpora, vocabulary levels, and text coverage, it is necessary to explore insights gained from the most prominent recipient of the fruits of corpus linguists’ labor: practitioners and researchers in the field of teaching English as a foreign or second language.
**Corpora: Utility and SLA applications**

The application of frequency data mined from corpora provides a number of advantages to both practitioners and researchers as opposed to other methods of linguistic data analysis. Defined simply as “a large, principled collection of natural texts”, corpora offer a glimpse into the “actual patterns of language use” (Reppen and Simpson: 2002, pp. 92-93). According to Biber, et. al (1998), four characteristics embody this approach: an empirical approach to observing natural language patterns, utilization of a “large and principled” group of texts as the base of analysis, extensive use of computers, and employment of both qualitative and quantitative techniques (p. 4). Aside from the utility of providing raw data regarding general frequency in a language, researchers such as Romer (2006) explore the validity of direct use of corpus materials inside the language-learning classroom. The goal of this approach is to “confront the learner as directly as possible with the data, and to make the learner a linguistic researcher” (Johns 2002, p.108, cited in Romer: 2006, p.124). The indirect approach instead is implicitly “researcher-focused” and oriented towards improving syllabus or materials design (Romer: 2006, p. 125). In the latter framework, students and teachers rely on researchers to filter and analyze corpus data on their behalf, and access the corpus data in an indirect manner. A great deal of studies have focused on methodology utilizing the latter approach, that is, analyzing textbooks and structural information presented in texts in light of frequency data.

**Textbook Studies, Corpora, and the Evaluation of “Vocabulary Levels”**

Early studies on textbook evaluation generally focused on appraising materials for purposes of use rather than analysis of specific features linguistic features (Williams: 1983;
Sheldon: 1988; Chambers: 1997). Even more recent studies by Ansary and Babaii (2002) eschew the former “checklist approach” supported by earlier studies, but still seek to establish “some theory-neutral, universal, broad characteristics” of EFL/ESL textbooks and establish criteria for evaluation based on ten EFL/ESL textbook reviews and ten EFL/ESL evaluations. Despite the strong tradition of utilizing textbook analysis to support adoption in curriculum, recent studies have focused moreso on micro features and linguistic structures within textbooks.

In reviewing the literature on textbook studies in the introduction to their own presentation of a corpus of English for General Purposes (EGP) textbooks, Meunier and Gouvernuer (2009) point to a focus on the “authenticity of input” provided by textbooks, as well as the prevalence of a focus on lexicon in recent textbook studies that utilized a corpus-based perspective. In the “authenticity” vein of research, Romer (2004a, 2004b, 2006b) analyzed the representation of modal auxiliaries, “if-clauses” in spoken language versus textbooks, and progressives, respectively in a variety of English as a Foreign Language Textbooks. Gilmore (2004) focused on discourse features of EFL and English for General Purposes Textbooks, and Hyland (1994) examined the use of modal verbs in English for Academic Purposes Writing Textbooks (cited in Meunier and Gouvernuer: 2009, p. 183). In the same line of research, Waugh and Fonseca-Greber (2002) found divergence in the usage of French subject clitics from a spoken corpus of Everyday Conversational Swiss French (ECSF) and Everyday Conversational Metropolitan French of France (ECMF) and the written forms traditionally presented in French as a Foreign Language textbooks.
Turning to the latter group of studies, or vocabulary-driven analyses, Nelson and West (2000) examined the divergence between unusually frequent and infrequent words presented in English for Business Purposes textbooks and a corpus of Business English in light of data from the British National Corpus (BNC). Biber, et al. (2004) examined the frequency of “lexical bundles” in comparing classroom and textbook presentations utilizing frequency data from the COBUILD Bank of English. Alternatively terming word groupings as “lexical phrases”, Koprowski (2005) analyzes these expressions as presented in three EFL/ESL textbooks by awarding “usefulness scores” to collocations, phrasal verbs, compounds, binomials, idioms, and fixed or semi-fixed expressions based on frequency and range and manual selection of data. Studies by Chujo and his colleagues apply a vocabulary levels and frequency framework in evaluating the comprehensibility of ESL textbook material utilizing a High Frequency Vocabulary Word List from the British National Corpus. Chujo and Nishigaki (2003) utilized the BNC Frequency List to measure vocabulary levels of EFL texts and the TOEFL, seeking to estimate coverage rates provided by textbook vocabulary for the TOEFL test and recommending a word list to cover words not found in textbooks. Chujo and Genung (2003) examined vocabulary levels of English for Special Purposes focused on Engineering and a variety of ESP texts, while Chujo (2004) delved more deeply into the application of vocabulary levels and frequency data as applicable to analysis of high school and college EFL textbooks and college qualification exams in the Japanese context. Studies by Chujo and his colleagues benefitted from the use of computer software such as TreeTagger, capable of transforming an unlemmatised or “raw” word list or text into a neatly lemmatized list organized by part-of-speech. In a context that does not
benefit from the same type of technological development, Lipinski (2010) and Godev (2009) utilized a page-by-page approach in comparing vocabulary from German and Spanish textbooks with respective corpora-derived frequency dictionaries of 5,000 and 6,000 words. Lipinski (2010) analyzed primary word lists presented in each chapter of three first-year German as a Foreign Language textbooks, while Godev (2009) provided estimates of coverage for randomly selected reading passages and coverage of the 1,000-5,000 word levels utilizing the end of the textbook vocabulary of one textbook and randomly selected chapter vocabulary lists. Godev’s (2009) study was limited by a lack of complete analysis of each textbook’s vocabulary, instead sampling “end-of-the-textbook vocabulary” of one book and vocabulary lists from randomly selected chapters. One limitation of these studies is the lack of consideration of which vocabulary items are most likely to be focused on, purposefully studied by students, and therefore acquired. A meta-analysis of L1 instructional vocabulary studies by Stahl and Fairbanks (1986) found direct vocabulary instruction using mixed methods including “drill and practice” strategies to encourage definitional knowledge of words as well as multiple exposures to vocabulary in context produced the largest effect sizes on eventual reading comprehension and measures of vocabulary. In light of these findings, Lipinski’s (2010) study may provide a sounder methodological basis, as primary chapter lists often contain vocabulary items students will be likely to learn for vocabulary quizzes and encounter throughout the chapter.

No similar study in the context of Teaching Arabic as a Foreign Language analyzes textbook contents utilizing corpora-derived frequency data. While a recent study by Belkouche, et. al (2010) created a corpus of primary school textbooks of language arts
textbooks for native Arabic speakers grades 1-6 in efforts to build a child-friendly Arabic dictionary, and efforts by Chik, et. al (2012) to establish a multidimensional framework for textbook analysis with Lahlali’s *Advanced Media Arabic* as the subject of study, there is little progress in efforts to evaluate and investigate the contents of books designed for TAFL in a methodologically rigorous or empirical fashion. The following study will attempt to address this gap in the literature by analyzing the most popular textbooks series for Teaching Arabic as a Foreign Language in America in light of corpus-based frequency data.
CHAPTER 3

METHODOLOGY AND SUBJECTS

Methodology

Taking a page from textbook studies utilizing corpora-derived data as a framework of analysis, this study will favor a “page-by-page” approach (Romer: 2004a; Gilmore: 2004; Hyland: 1994; Koprowski: 2005; Godev: 2009; Lipinski: 2010) over a corpus-based methodology. The former is preferred over the latter, in which the sum of all the language presented within textbooks is privileged over saliency of specific vocabulary items (Nelson and West: 2000; Biber, et. al: 2004; Meunier and Gouvernour: 2009). This study will first alphabetize primary chapter vocabulary lists, “Remember This Vocabulary”, and “Learn This Vocabulary” sections presented in Part One of the most popular textbook series for Teaching Arabic as a Foreign Language, *Al-Kitaab fii Ta’allum Al-‘Arabiyya* (Brustad, et. al: 2004). Despite the appeal of the “pedagogical corpus”-based approach, where analyzed data is extracted from a constructed textbook corpus based on frequency within the textbook (Meunier and Gouvernuer: 2008), the “page-by-page” approach brings focus to vocabulary that will most likely be taught or studied, focused on throughout the chapter, and therefore is the vocabulary most likely to be acquired (Stahl and Fairbanks: 1986).

As evidence points to the importance of the first 3,000 word families or 5,000 discreet words as the “turning point” of vocabulary knowledge for reading (Laufer: 1992), with the first 1,000 word families or 1,600 words covering 77% of written texts in English (Nation: 2001), this study will ultimately examine what kind of minimal receptive vocabulary knowledge for reading could be obtained after completing Book One of the *Al-Kitaab* series.
Part one of the study will propose minimal word families from memorized textbook vocabulary from the primary (“New Words” section) and secondary (“Remember These Words”) vocabulary lists. As learners may conceptualize word knowledge based on the consonantal root (Ravid: 2005) and memorize items without recourse to the root and pattern system (Jiang: 2004), items without transparent consonantal roots that could be easily identified by novice learners will be similarly alphabetized. While data on the role of the consonantal root and word/stem is not found within the AFL context, it is possible to base Arabic L2 word families on a hypothesized organization of the mental lexicon after evidence from English proposing psychological reality of word families (Nagy, et.al: 1989; Bertram, Laine & Virkkala, 2000, cited in Nation: 2012). The hypothesized root and word-based families will be analyzed according to lemmas, lexical items, and minimal word families without recourse to the morphological patterns that could extend receptive vocabulary knowledge. Part one of the study will be guided by three primary questions:

1) How many total lexemes (including MWUs) could a student recognize receptively if the primary vocabulary from Part One of the series was memorized?

2) How many lemmas (excluding MWUs) could a student recognize receptively if the primary vocabulary from Part One of the series was memorized?

3) On average, how many family members do the minimal word families contain?

Hypothetical word families will be based on Bauer and Nation’s (1993) definition of word families as a “base word and all of its derived forms that can be understood without having to learn each form separately” (p. 253). Initial counts of the total number of lexemes (Schmitt: 2000, p. 2) will consider not only discrete words but also “multi-word units.” This measure will account for items like al?umam al-muta?hida (“United Nations”) containing
multiple tokens, as these items are more likely to be “learned, stored, and used as complete units” (Nation and Meara: 2002, p. 36). It will also include the arithmetic mean ($\bar{x}$) of the number of members per family in these hypothetical word families after Nation (1983, cited in Bauer: 1997, p. 24). This section will also include observations regarding observed patterns within word families.

Part two of the study will examine these word families in light corpus-derived frequency data up to the 5,000-word level from *A Frequency Dictionary of Arabic* (Buckwalter and Parkinson: 2011). This corpus-derived dictionary meets the first of Biber, et. al’s (1998) requirements for a corpus as a representative sample of natural language, as the dictionary was culled from a corpus of 30 million tokens of 90% written and 10% spoken texts (Buckwalter and Parkinson: 2011, p. 3). Buckwalter and Parkinson’s (2011) efforts to present final frequency counts adjusted for range or dispersion across each text type within the corpus strengthens the *Dictionary* as a tool to examine the “actual patterns of language use” within modern Arabic (Reppen and Simpson: 2002, pp. 92-93).

In this way, lexical items and word families can be evaluated according to data culled from a sizeable corpus from multiple genres, pointing to a comparison between textbook vocabulary and which words are most likely to be encountered by students, and are therefore most useful (Nation and Waring: 1997). Since comparisons between proposed vocabulary levels and text coverage in English (Nation: 2006) and Arabic (Van Mol: 2006) reveal a reader of Arabic news texts must know at least 2,100 more discreet words than an English reader and approximately 5,600 more words to access a novel at a 95% coverage rate, estimations for more elementary levels of text coverage may be higher than English. It then
seems likely that Van Mol’s (2000) estimation of 3,000 words as a Breakthrough Basic level and Al-Batal’s (2006) estimation of 3,000-3,500 words as the minimum vocabulary amount for advanced proficiency may not yield the same proposed rate of text coverage as in English. The number for minimal text coverage may instead lie between Nation’s (1990) estimate of 1,600 words (1,000 word families) and Laufer’s (1992) estimate of 5,000 words (3,000 word families).

Part two of the study will examine these minimal word families in light of data from the *Frequency Dictionary of Arabic* (Buckwalter and Parkinson: 2011) and will be guided by the following questions:

1) What kinds of coverage rates of the 1,000-5,000 word levels within the *Dictionary* can be achieved using textbook vocabulary?

2) What kinds of patterns of divergence emerge when analyzing textbook vocabulary in light of frequency data?

Coverage rates will be assessed by calculating the number of lemmatized vocabulary items found at the 1,000, 2,000, 3,000, 4,000, and 5,000 word level in the same vein as previous textbook studies utilizing a word-levels approach to vocabulary analysis (Chujo and Nishigaki: 2003; Chujo and Genung: 2003; Chujo: 2004; Godev: 2009; Lipinski: 2010). Due to indications of the importance of the 3,000-3,500 word levels in Arabic (Van Mol: 2000; Al-Batal: 2006), analysis will be carried out for the first 1,000-3,000 words in 100-word increments. Coverage rates at each level will also be compared. Analysis of reoccurring patterns of divergence in terms of discontinuous or “broken” plurals and verb patterns between the two data sets will also be carried out to inform actionable pedagogical recommendations outside simply targeting word level increments or levels with the least
amount of coverage. Pedagogical recommendations and implications of findings, including direct instruction of high frequency verbs at the basic word levels (1,000-3,000) with the largest gaps, introduction of frequent but alternative forms for nouns and verbs present in the textbook but not the Frequency Dictionary, and extensive reading (Day and Bamford: 2002) at the advanced level will follow a discussion of the results. Testing theoretical word families to measure the vocabulary knowledge of native and non-native Arabic speakers and the relationship between assessments of vocabulary and reading proficiency will also be discussed as directions for future research into decreasing the “vocabulary burden” for learners of Arabic as a Foreign Language (Nation: 2001, cited in Young: 2011).

This study will also venture into new territory in addressing Nation’s (2001) three questions utilized to define vocabulary learning goals: how many words are there in the language, how many words do native speakers know, and how many words are necessary to carry out tasks using the language. As there is little data supporting the exploration of the first two questions in the context of acquisition of Arabic by native or non-native speakers, this study represents a new way of examining these larger questions by asking what kind of lexical knowledge a non-native speaker may have after completing Part One of the most popular Arabic as a Foreign Language textbook series.

This study suffers from a few limitations. Since word families in Arabic remain terra incognita, proposed word families may not necessarily reflect the exact number or type of words a student knows receptively, as a test of word knowledge will not be conducted. In reality, a learner’s mental lexicon may contain more than these minimal word families, as different Arabic language programs may cover more or less of the vocabulary in this study,
especially as vocabulary from the Spoken Egyptian Arabic section of each chapter was not analyzed. Students may also have more or less control over the inflectional and derivational processes presented in the textbook series, which would extend or diminish theoretical receptive knowledge. There may also be gaps between the vocabulary analyzed and the vocabulary actually acquired by students, as specific verbs, function words, content (lexical) words, and expressions presented formally will not be analyzed if these items appear outside primary and secondary vocabulary lists. Lastly, since frequency data closer to the minimal level of 8,500 words for minimal coverage of a newspaper text was not available, analysis of results from proposed word families will revolve around estimations of 3,000 and 5,000 word levels as minimums for text coverage and general linguistic competence in Arabic (Van Mol: 2001, 2006; Al-Batal: 2006). While it is not possible to separate frequency data derived from written texts within the *Frequency Dictionary*, frequency data may be thought to represent mostly written data, as 90% of data sources were from written texts with only one corpus (Internet discussion forums) being likely to contain a great deal of spoken Arabic lexicon. With the exclusion of postings on Internet discussion forums, it is probable that the remaining texts types (newswire, editorials, and learned prose) will be of the same variety as the vocabulary analyzed, or mostly MSA.

*The textbook: The Al-Kitaab Fii Ta'allum al-'Arabiyya, Part One*

*Pedagogical philosophy and structure*

The *Al-Kitaab* series (Brustad, et. al: 2004) is one of the most popular TAFL textbook series used in American universities and study abroad programs, and is recommended by the National Middle East Learning Resource Center (NMERLC: ND). Part One is designed to
follow *Alif Baa* (Brustad, et. al: 2004), which introduces Arabic language-specific orthography and phonology. Part One of the series covers 150 instructional hours and aims to advance skills in reading, writing, listening, and speaking in MSA to the intermediate level (Brustad, et. al: 2004a, p. xvi). The narrative texts in Part One are oral recordings and follow the story of two young Egyptians, Maha and Khalid. Stories from their daily lives and the lives of their family members range from relative simplicity (Chapter One: *ʔana Maha*，“I am Maha”) to greater levels of conceptual complexity (Chapter 12: *ʔasaSab qara:r fi: haya:ti*; “The most difficult decision in my life”; Chapter 19: *ma:θa yaqsuda:n bikalima muna:sib?!*, What do they mean by the word “appropriate”?). The series is designed after the communicative approach to Arabic instruction. The “narrative thread” at the beginning of each chapter is presented in a “simplified *fuṣha*” while dialogues between the characters are presented in spoken Arabic, as “any dialogue in *fuṣha* between the characters of the story will be contrived” (xix, Arabic Introduction). Characters presenting themselves in a monologue in video recordings at the beginning of each chapter represent a linguistic reality in which “formal or quasi-formal” speeches are delivered in MSA (xx). Spoken Arabic (*ʕami:yya*) is introduced to students “gradually” through the dialogues, exposing students to the “sounds of *‘amiyya* and its vocabulary” (xx).

**Vocabulary**

The first portion of each chapter prioritizes the “acquisition and activation of vocabulary” (xi). Vocabulary present in monologues by Maha, Khalid, or their family members appears in a list at the outset of each chapter and is recorded on the accompanying CD. CD recordings contextualize vocabulary items through presentation in complete
sentences. Vocabulary is recycled in a variety of exercises designed to enhance reading, writing, speaking, and reading skills. According to the authors’ pedagogical philosophy, vocabulary acquisition is the “core of building proficiency in Arabic” (xi). They argue success of a language program is based on what a student can do with the language outside the classroom (xxi). The employed method of vocabulary instruction offers students the “opportunity…to accumulate knowledge and vocabulary”, arguing vocabulary lists may be useful for review but not as a substitute for the gradual accumulation of vocabulary knowledge (xxi). For vocabulary key to understanding texts and exercises, new vocabulary items are introduced before the exercise in question under a section entitled “Learn These Words.” A second section at the end of each chapter titled “Remember these words” reinforces important vocabulary encountered throughout the chapter and repeats some of the vocabulary items presented in the “Learn these words” section.

Figure 3.1-Example of “Learn These Words” from Part One (Brustad, et. al: 2004, p. 78)
The authors introduce derived and inflected forms gradually throughout Book One, in keeping with the theoretical basis of vocabulary as an “accumulation” of knowledge (p. xxi). It represents a gradual approach to introducing inflectional and derivational morphology, a method that is continued throughout the second book where the size of word lists is doubled, with new words “From the Dictionary” and a second section of vocabulary derived from the same root known as “Study and Learn.” While derivational morphology is not formally introduced in Book One until page 250, including a full explanation of each of the main ten verb forms and their semantic associations, the important concept of relatedness among word containing the same root is introduced on page 133. Many exercises also include guessing roots from derived forms. In this respect, when the basic tools for Arabic derivational morphology are introduced, the core principle of derivational morphology, that is, semantic relatedness of derived forms and roots, is not entirely new. Moreover, the presentation of new verbs along with the derived present tense verb and verbal noun as early as Chapter 9 (p. 148) encourages the building of small word families even before the complete version of verb morphology is introduced. After the introduction of the most common derived verb
forms, later chapters offer more detailed explanations of the semantic and morphological relations between verb forms, continuing the discussion in Part Two of the textbook series.

Figure 3.3—Primary vocabulary list before introduction of derivational morphological processes (Brustad, et. al: 2004, p. 35)

Figure 3.4—Limited word families in primary vocabulary after presentation of some derivational and inflectional morphology (Brustad, et. al: 2004a, p. 148)
The frequency data: A Frequency Dictionary of Arabic

*A Frequency Dictionary of Arabic: Core Vocabulary for Learners* (Buckwalter and Parkinson: 2011) is a pedagogical resource intended to “prepare students of Arabic for the words they are most likely to encounter in the real world”, including data from both Spoken and Standard Arabic (p. 1). The lexical data represents the most frequent Arabic words from a corpus of 30 million tokens culled from a variety of sources. Ten percent (10%) of the data was gathered from transcriptions of “unscripted” speech from Egyptian, Levantine, Iraqi, Gulf, and Algerian spoken varieties, while the latter 90% of the data was derived from various corpora of written texts. For written texts, researchers first culled raw frequency data from five different text types—daily news, newspaper editorials, opinion essays, and regular newspaper columns, “learned prose” extracted from scholarly journals, posting on Internet discussion forums, and literature from short stories, novels, and plays. Each corpus contained sources from multiple Arabic-speaking locales, and “practically all” of the texts were published in the period of 2006-2007 (Buckwalter and Parkinson: 2011, p. 3). As a result, the Dictionary represents a snapshot of the Arabic language across the Arabic-speaking world at the time of compilation.

<table>
<thead>
<tr>
<th>Text Type</th>
<th>Source</th>
<th>Number of documents and tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newswire section</td>
<td>Official news agency of each country from 2006; front page news items for national newspapers from 2007</td>
<td>27 documents, 5.4 million tokens</td>
</tr>
<tr>
<td>Newspaper editorials, opinion essays, and regular columns</td>
<td>“Long-standing and well-known” Arabic newspapers and leading pan-Arab newspapers</td>
<td>27 documents, 5.4 million tokens</td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
<td>Documents</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>“Learned prose”</strong></td>
<td>Academic, scientific, religious, and popular formal writing; weekly magazines and monthly journals, “Friday supplement” from newspapers</td>
<td>27</td>
</tr>
<tr>
<td><strong>Internet discussion forum postings</strong></td>
<td>Arabic Internet discussion forums on youth culture, concerns of women, religious issues, medical advice, personal and family issues</td>
<td>27</td>
</tr>
<tr>
<td><strong>Literature</strong></td>
<td>Works by Naguib Mahfouz, Tawfiq Al-Hakim, Ali Salim, Edwar Al-Kharrat, Tayeb Salih, Ghassan Khanafani, Elias Khoury, Khalil Gibran, Ahlan Mosteghanemi, Tahar Ouettar, Najwa Barakat, Rajaa Alsanea, and Alaa Al-Aswany</td>
<td>72</td>
</tr>
</tbody>
</table>

The *Dictionary* contains a main frequency list in descending order as well as two glossaries organized by function and part of speech. Words in the final frequency list from one to 5,000 were adjusted for frequency across each text type or “range.” For example, *qarrar* (“to decide”) had a raw frequency of 5019 within the raw list but the dispersion rate or “range” was 0.9660, leading to an adjusted frequency rate of 4848.6 and a final ranking of 767 (Buckwalter and Parkinson: 2011, p. 6). Listings within the dictionary contain information regarding the frequency rank, the headword (primary uninflected word), part(s) of speech, most common English glosses, a sample sentence from the corpus and English translation, dispersion across text types (range), raw frequency, and register variation indicating whether the item occurs in the literature, news, spoken, or online forums corpus (Buckwalter and Parkinson: 2011, p. 6-7). Plurals of nouns are also listed beside the headword. Listed plurals are attested in the corpus. Thirty thematic “call-out boxes” are also
interspersed throughout the frequency list, including topics such as Animals (p. 9), the Body (p. 16), Food (p. 23), Clothing (p. 30), Transportation (p. 36-37), Family (p. 44-45), Materials (p. 51), and other topics. The boxes include topical vocabulary organized according to frequency as well as data above the 5,000 word level even if the item is not found in the main 5,000-item frequency list.
CHAPTER FOUR
RESULTS, DISCUSSION, PEDAGOGICAL RECOMMENDATIONS, AND CONCLUSION

Minimal word families from Al-Kitaab fii Ta'allum Al-‘Arabiyya Part One

Duplicate items were removed from the lists before alphabetization. Words without transparent roots and borrowed words were fully listed without recourse to the consonantal root. Although an advanced learner may recognize the shared root of \( w-h-d \) in \( wa\hda \) (“loneliness”) and \( mutahida \) (“united”), a novice learner may instead rely on a direct translation of the L1 concepts and therefore memorize certain lemmas as single units unless direct attention is drawn to the root contained within the word. Multi-word units, such as \( al-\h namedal-mutahida \) (“United Nations”) and prepositions or function words such as \( wa \) (“and”) or \( ?amma...fa \) (“as for”) were similarly alphabetized. Multi-word units were not separated into disparate words as students at the projected level (Novice High/Intermediate Low) may analyze these phrases as chunks instead of two separate words. The remaining vocabulary items derived from primary vocabulary lists, secondary vocabulary lists (“Remember These Words”), and other lists containing instructions encouraging students to memorize were alphabetized according to triconsonantal root. Word families were counted as the base, derived, and inflected forms. Thus, all words under the lemma \( tarak \) (“to leave, abandon”), including \( yatruk \) (3PMSG present imperfect verb, “he leaves”) and \( tark \) (verbal noun; “leaving, abandoning”) were included as members of the same word family, leading to the total number of family members for word family one under the root \( t-r-k \) as three. The total number of disparate lexemes and lemmas were counted by hand. In counting lemmas, only the citation form of the lemma, or the third person masculine singular past perfect verb,
was counted. The third person masculine singular present imperfect verb was considered part of the same lemma. For example, a word family from the verb *daxxan* (“to smoke”) also contained *yudaxxin* (“he smokes”) as well as the verbal noun *tadxi:n* (“smoking”). However, only *daxxan* and *tadxi:n* count as lemmas, as *yudaxxin* is an inflected form of the base *daxxan*. As multi-word units (MWUs) were not counted, a second count was taken to include these vocabulary items. Lexemes were counted in the same fashion as lemmas, but included MWUs like *min ?aṣl* (“of ____ descent”) and *fi ra?hi* (“in his opinion). The arithmetic mean (\( \bar{x} \)) was also calculated for the number of word families per root and number of number of words per family.

**Figure 4.1-Totals and means from Part One Word Families**

<table>
<thead>
<tr>
<th>Total number of lexemes (including MWUs)</th>
<th>Total number of lemmas (excluding MWUs)</th>
<th>Mean number of members per family (( \bar{x} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>697</td>
<td>648</td>
<td>2.07</td>
</tr>
</tbody>
</table>

Results for the mean number of words per family, modeled after Nation’s (1983, cited in Laufer: 1997, p. 24) calculation in English, showed basic families assembled from the vocabulary presented in Book One of the *Al-Kitaab* series without recourse to derivational processes contain an average of 2.07 words. Vocabulary showed a great deal of variation among size and type of word families, especially in light of some items with non-productive roots. Some roots had greater numbers of vocabulary, but due to the incremental nature of word family development as proficiency increases (Bauer and Nation: 1993) items that may be considered part of the same family for advanced learners may separate into separate families for students at lower levels of proficiency who only completed Book One of the
series. At this level, separation may be hypothesized based on memorized meanings in English associated with Arabic due to a lack of complete control over derivational morphology and semantic associations of each derivational form. Moreover, some words may be accessed according to the first consonant within the word if the underlying morphological patterns of derivation are not understood. If there are not many other lemmas or lexical items sharing the same root, or if the root consonant is not clear within the word, students may be likely to rely on memorized meanings associated with each word rather than the consonantal root. It seems unlikely students will know the underlying root of both ʔibn (son) and bint (girl) is b-n. On the contrary, some roots showed a greater number of easily relatable family members, even if the student is hypothetically only memorizing vocabulary and not using Arabic’s derivational processes. Roots like k-t-b (write) possess a more transparent relationship between family members and therefore a larger number of members.

Figure 4.2-Word family for k-t-b (write) from Al-Kitaab Part One

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Transcription</th>
<th>Description</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>كتاب</td>
<td>katab</td>
<td>3FMSG, Past Perfect, Pattern I</td>
<td>wrote</td>
</tr>
<tr>
<td>كتاب</td>
<td>kita:b</td>
<td>Derived Noun</td>
<td>book</td>
</tr>
<tr>
<td>مكتبة</td>
<td>maktab</td>
<td>Noun of place, Pattern I</td>
<td>office</td>
</tr>
<tr>
<td>مكتبة</td>
<td>maka:tib</td>
<td>noun of place, plural, pattern I</td>
<td>offices</td>
</tr>
<tr>
<td>مكتبة</td>
<td>maktaba</td>
<td>Noun of place, Pattern I</td>
<td>library</td>
</tr>
<tr>
<td>مكتبة</td>
<td>maka:bat</td>
<td>noun of place, pattern I plural</td>
<td>libraries</td>
</tr>
<tr>
<td>كتاب</td>
<td>yaktub</td>
<td>3FMSG Pres Imp</td>
<td>he/its writes</td>
</tr>
<tr>
<td>كتاب</td>
<td>alkitaba</td>
<td>Verbal Noun, Pattern I</td>
<td>writing</td>
</tr>
<tr>
<td>كتاب</td>
<td>kastib</td>
<td>MSG Active Participle, Pattern I</td>
<td>writer</td>
</tr>
<tr>
<td>كتاب</td>
<td>kutab</td>
<td>MPL Active Participle</td>
<td>writers</td>
</tr>
</tbody>
</table>
Other lemmas did not break down into a number of other forms, such as *birnaːmij*, *biraːmij* (“program”, “programs”). Secondly, function words like prepositions and the definite article (*al*-) are not necessarily productive from a word families’ perspective but may be useful if found relatively early in the frequency list. These words may be likely to be encountered in written texts despite lack of a number of related forms and relatively small family size.

*Part Two: Evaluation of vocabulary based on frequency data*

After vocabulary items were lemmatized and counted according to roots, lexemes, lemmas, word families, and average number of members per root and family, lemmatized families were examined in light of data from the *Frequency Dictionary of Arabic* (Buckwalter and Parkinson: 2011). Each item in the list of lemmatized vocabulary was first tagged according to its entry in the corpus-derived frequency dictionary.

*1-999 Word level*

In analyzing tagged vocabulary from the aforementioned textbook lists, coverage of words between the most frequent word according to the corpus-derived data (*al*, definite article, “the”) to the 999th most common word (*mustamirr*, past participle, “continuous, incessant; continuing”) revealed the following levels of coverage. Results are divided into 100-word increments.
Figure 4.3-Coverage rates of textbook vocabulary, 1-999 word level

The combined amount of coverage for textbook vocabulary of the first 1,000 most frequent words in the corpus-derived data reveal the greatest coverage levels occur at the 0-99 (49%), 101-199 (51%), 200-299 (39%), and 400-499 and 500-599 levels (34%). Overall coverage of lexemes from textbook vocabulary totaled to 317 of the total 1000 possible items, or 31.7%.

1000-1999 Word level

Results at the 1000 word level, ranging from the 1000th most frequent word (kida, adverb, “thus, this way, like this”) to the 1999th most frequent word (tafsi:r, derived noun, “explanation, commentary, Quranic exegesis”) showed the following levels of coverage.
The greatest coverage rates appear to cluster at the 1100-1199, 1300-1399, and 1500-1599 word levels. Combined totals of 100-word increments reveal a combined coverage rate of 123 out of the 1000 total items, or 12.3% coverage.

2000-2999 Word level

Coverage rates at the 2000 word level between the 2000th most frequent word (janaḥ, noun, “wing”) and the 2999th most frequent word (majbax, noun of place, “kitchen”) for textbook vocabulary show the following levels of coverage.
Greatest coverage rates occurred at the 2000-2099 word level (7%) and 2900-2999 level (8%). Combined coverage rates for the level amounted to 58 words out of 1000, or 5.8%.

**3000-3999 and 4000-5000 Word Levels**

Coverage rates were next tabulated at the 3000 word level spanning the most frequent words between the 3000\textsuperscript{th} most common word in the corpus-derived data (\textit{zirq}, noun, livelihood, sustenance; daily bread) to the 3999\textsuperscript{th} most frequent item (\textit{niqa:b}, noun, “full faced veil; niqab). The greatest coverage rates occur at the 3000-3099 (5%), 3400-3499 (5%), and 3500-3599 (5%) levels. Overall, coverage rates of this level yield coverage of 23 words out of 1000, or 2.3% of the total.

Rates of coverage for words at the 4000 word level, between the 4000\textsuperscript{th} most frequent word across the various corpora (\textit{munfaṣal}, past participle/adjective, “separate, disconnected;
detached, loose”) and the 5000th most common word in the corpora-derived data (damm, noun, “joining, addition, annexation”) revealed an overall coverage rate of 19 out of 1000 items, or 1.9% of the total at this level. The highest coverage rate was found at the 4100-4199 level at 7%.

**Combined totals**

The coverage rates across the 0-999, 1000-1999, 2000-2999, 3000-3999, and 4000-5000 word levels revealed the greatest coverage levels reside at the 0-999 and 1000-1999 word level, at 31.8% and 12.3% respectively. Lower rates of coverage exist at the higher levels of 2,000-3,000 (5.8%), 3000-3,999 (2.5%), and 4000-5000 (1.9%).

Figure 4.6-Overall coverage rates of textbook vocabulary, 1-5,000 word levels

**Patterns of divergence between the textbook vocabulary and corpora-derived data**

The process of tagging textbook vocabulary to estimate coverage rates of the different vocabulary levels revealed patterns of divergence between “broken” plural nouns and
patterns for verbs presented within the textbook and patterns found within the corpus-derived data.

“Broken” Plurals

Broken plurals of some nouns within the textbook vocabulary were not always found within the frequency data, as they were not attested within the corpus of 30 million tokens.

Figure 4.7-Singular nouns with plurals unattested in the Frequency Dictionary

<table>
<thead>
<tr>
<th>Singular Noun</th>
<th>Plural listed in Al-Kitaab</th>
<th>Data from FD</th>
<th>Gloss/Translation</th>
<th>Rank of FD</th>
<th>Singular Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>جامع</td>
<td>جامع</td>
<td>تجوامع</td>
<td>جامع &quot;mosque&quot;</td>
<td>2812</td>
<td></td>
</tr>
<tr>
<td>حداء</td>
<td>حداء</td>
<td>حديداء</td>
<td>حداء &quot;shoe/pair of shoes&quot;</td>
<td>3637</td>
<td></td>
</tr>
<tr>
<td>خال</td>
<td>خال</td>
<td>خال</td>
<td>خال &quot;maternal uncle&quot;</td>
<td>3188</td>
<td></td>
</tr>
<tr>
<td>دفتر</td>
<td>دفتر</td>
<td>دفتر</td>
<td>دفتر &quot;notebook&quot;</td>
<td>3904</td>
<td></td>
</tr>
<tr>
<td>شكاك</td>
<td>شباك</td>
<td>شباك</td>
<td>شباك &quot;window&quot;</td>
<td>4148</td>
<td></td>
</tr>
<tr>
<td>شقة</td>
<td>شقة</td>
<td>شقة</td>
<td>شقة &quot;apartment&quot;</td>
<td>2321</td>
<td></td>
</tr>
<tr>
<td>طابق</td>
<td>طابق</td>
<td>طابق</td>
<td>طابق &quot;floor, story&quot;</td>
<td>3892</td>
<td></td>
</tr>
<tr>
<td>عطلة</td>
<td>عطلة</td>
<td>عطلة</td>
<td>عطلة &quot;vacation, holiday, recess&quot;</td>
<td>2947</td>
<td></td>
</tr>
<tr>
<td>مطافخ</td>
<td>مطافخ</td>
<td>مطافخ</td>
<td>مطافخ &quot;kitchen&quot;</td>
<td>2999</td>
<td></td>
</tr>
<tr>
<td>جذ</td>
<td>جذ</td>
<td>جذ</td>
<td>جذ &quot;grandfather&quot;</td>
<td>1226</td>
<td></td>
</tr>
</tbody>
</table>

In these cases, the singular noun presented in the frequency dictionary and the textbook was identical, but the broken plural was not listed. Other plurals present within the frequency data displayed additional types of divergence: alternative broken plurals, additional entries for broken plurals beside the headword, or broken plurals listed as plurals for different headwords.
Alternative and Additional Broken Plurals

While the previous data set included singular nouns presented with specific broken plurals within the textbook that were unattested in the frequency-derived data, other cases displayed alternative or additional broken plurals beside the singular headword. In the first case, a number of broken plurals found within the textbook vocabulary were not attested within the frequency dictionary, once again implying they were not attested within the 30 million token corpus.

Figure 4.8-Alternative Broken Plurals

<table>
<thead>
<tr>
<th>Singular Noun</th>
<th>Al-Kitaab</th>
<th>Frequency Dictionary</th>
<th>Gloss/Transliteration</th>
<th>Rank of FD Broken Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>ابناء</td>
<td>بنو</td>
<td>أبناء</td>
<td>بنو</td>
<td>&quot;sons, children&quot;</td>
</tr>
<tr>
<td>عموم</td>
<td>عام</td>
<td>عموم</td>
<td>عموم</td>
<td>&quot;paternal uncles&quot;</td>
</tr>
<tr>
<td>قمصان</td>
<td>قمص</td>
<td>قمصان</td>
<td>قمصان</td>
<td>&quot;shirts&quot;</td>
</tr>
<tr>
<td>نادي</td>
<td>نادي</td>
<td>نادي</td>
<td>نادي</td>
<td>&quot;clubs, associations&quot;</td>
</tr>
<tr>
<td>ورق</td>
<td>ورق</td>
<td>ورق</td>
<td>ورق</td>
<td>&quot;papers&quot; (coll. noun)</td>
</tr>
</tbody>
</table>

Some headwords did contain a listing for the broken plural listed in the textbook, but also sometimes contained an extra broken plural within the Dictionary.

Figure 4.9-Additional Broken Plurals

<table>
<thead>
<tr>
<th>Singular Noun</th>
<th>Al-Kitaab</th>
<th>Additional Form in Frequency Dictionary</th>
<th>Gloss/Transliteration</th>
<th>Rank of Broken Plurals</th>
</tr>
</thead>
<tbody>
<tr>
<td>أغنيات</td>
<td>أغاني</td>
<td>أغنيات</td>
<td>أغاني</td>
<td>&quot;songs&quot;</td>
</tr>
<tr>
<td>طلاب</td>
<td>طالب</td>
<td>طلاب</td>
<td>طالب</td>
<td>&quot;students&quot;</td>
</tr>
<tr>
<td>مرات</td>
<td>مرة</td>
<td>مرات</td>
<td>مرة</td>
<td>&quot;times, moments, occasions&quot;</td>
</tr>
<tr>
<td>أماكن</td>
<td>مكان</td>
<td>أماكن</td>
<td>مكان</td>
<td>&quot;places&quot;</td>
</tr>
</tbody>
</table>
**Divergent Broken Plurals and Singular Nouns**

Some singular noun and plural pairs listed within the textbook did not necessarily match the forms in the textbook and sometimes broken plurals from a specific pair were listed as plurals of other singular nouns of the same root. However, this was only the case for three specific noun/plural pairs: *fikra* (“thought”), *mawṣīd* (“appointment; deadline”), and *qari:b* (“relative”).

**Figure 4.10-Divergent singular/plural noun pairings**

<table>
<thead>
<tr>
<th>Al-Kitaab Singular-Plural Pair</th>
<th>Frequency Dictionary Singular-Plural Pairs</th>
<th>Gloss/Transliteration</th>
<th>Frequency Rank of FD Pairing</th>
</tr>
</thead>
<tbody>
<tr>
<td>فكرة أفكار, فكرة أفكار</td>
<td>ِfikr, ِfikr, &quot;thinking, idea, concept&quot;; ِfikra, &quot;idea, concept, notion&quot;</td>
<td>Pair One: 522, Word Two: 471</td>
<td></td>
</tr>
<tr>
<td>قريب أفكار, قريب أفكار</td>
<td>qari:b, qara:ʔib, &quot;near, related; relatives&quot;; ʔaqrab, ʔaqrab, &quot;near/nearest&quot; (elative), &quot;relatives&quot;</td>
<td>Pair One: 476, Pair Two: 1354</td>
<td></td>
</tr>
<tr>
<td>موعد مواعيد, موعد مواعيد</td>
<td>mawṣīd, mawaṣi:d, &quot;appointed time, deadline&quot;; miːʕa:d, mawaṣi:d, &quot;appointment, promise&quot;</td>
<td>Pair One: 1005; Pair Two:3315</td>
<td></td>
</tr>
</tbody>
</table>

In some cases, there appears to be a slight difference in semantic meaning between some singular-plural pairs presented in the textbook and in the frequency dictionary. In the instance, there may be a slight difference in meaning between *mawṣīd* (“appointed time, deadline”) and *miːʕa:d* (“appointment, promise”). A final form of divergence emerged in analyzing the variances between some verbs or verbal nouns presented within the textbook and forms under the same root with the same general root attested within the corpus-derived frequency data.
Divergent verbs, verbal nouns, and other forms

Some verbs and verbal nouns, while contained within the textbook vocabulary, were not attested in precisely the same form within the frequency data. Some verbal nouns such as ʔizdiḥaː:m (“crowdedness”) were not attested in the frequency data, but a different form with a nearly identical semantic meaning and corresponding listing under the same root (zahma) was attested as the 4048th most frequent word. Similarly, the verbal noun muḍaːkira (“studying, reviewing lessons”) was not attested within the frequency data and a suitable alternative with the same semantic association under the same root was not found. Some verbs like inšaqal (“to be occupied with s.th.”) were not attested in the frequency data and a verb with similar semantic associations under the same root could not be obtained. A similar pattern emerged in observing relationships between some verbs within the textbook vocabulary and the frequency data, but a semantically related verb of the same root. The intransitive verb taḡayyib (“to be absent from”) was not attested among the 5,000 most frequent words in the dictionary, but a secondary form presented within the textbook (ḡaːb, “to be or remain absent”) was attested at the 1000-word level at number 1607. For some verbs, a verbal noun was attested while the past tense verb was not found in the corpus. This may be due to language change, by which a verb like darras (“to teach”, literally “to make someone study”) was only attested within the corpus as a verbal noun tadriːs (“teaching, instruction, pedagogy”; number 4808). Instead, a synonym ʕallam (to teach, literally to make someone know; number 3450) was also attested as a verbal noun, taʃliːm (“teaching, pedagogy”, number 388). This may indicate the lemma ʕallam is more productive and more likely to be encountered. Lastly, some vocabulary items such as mašuːr (“famous, famed”)
were attested in the dictionary, but a similar form under the same root with similar semantic associations (šahi:r, “famed, renowned”) was listed at a higher frequency level (2563 versus 2142).

Discussion of results

This study represents an attempt to catalogue the lexical contents of the most popular TAFL textbook as well as to question how much of this vocabulary is found within a comprehensive Arabic corpus of 30 million tokens, and is therefore likely to be met by students during reading. Organizing the contents of the textbook utilizing a word families framework gives credence to anecdotal accounts of a greater “learning burden” reported by Young (2011) specifically for Arabic. In acquisition of Arabic vocabulary, greater amount of effort must be put forth to acquire a greater number of lexemes in order to build word families and utilize the language. While projections for an average number of words per family applies only to vocabulary lists found in Part One of the Al-Kitaab series, estimates even without recourse to derivational and inflectional patterns that could extend the size of students’ word families is higher than Nation’s (1983) projected average of 1.60 words per family for English word families as a whole. With an average of 2.07 members per family, a complete projection of word family size for learners of Arabic as a Foreign Language will be much larger and increase as the student acquires greater control over the derivational and inflectional morphological patterns of the language, as suggested by Bauer and Nation (1993). Unfortunately, these word families remain purely theoretical, as they were not tested on students to measure vocabulary knowledge and were not extended in light of morphological processes presented within the textbook. Based on the size of word families
within the Arabic Foreign Language context, it could also be hypothesized that word families within the Arabic language as a whole are much larger than English. While these results may only apply to AFL students, analysis of very limited AFL vocabulary word families offers a glimpse into the vastness of word families in the Arabic language at large.

Part two of the study analyzed the quality of textbook vocabulary utilizing coverage rates of basic textbook vocabulary extracted from primary and secondary vocabulary lists of Part One of the *Al-Kitaab* textbook series. Textbook items were then compared to the 5,000 most frequent lemmas attested within a corpus compiled from a representative sample of spoken and written language. Examining textbook vocabulary in light of a sizeable and representative sample of the language points to the kind of vocabulary in contemporary usage and the degree of overlap between these frequent words and textbook vocabulary. The overlap or divergence between lemmas within the *Dictionary* and textbook vocabulary represents the relationship between the language commonly utilized by Arabic speech communities and the type of language acquired by non-native learners of Arabic as a Foreign Language. While projections for coverage of these vocabulary levels presents a somewhat bleak picture of the relationship between textbook vocabulary and actual language use in written and spoken contexts, results should be analyzed with caution. As this study was primarily interested in the basic contents of the textbook, every single vocabulary item presented within the textbook was not analyzed. High-frequency content and function words presented within grammar sections or activities may have been excluded by virtue of focusing only on the primary vocabulary lists. Secondly, many proper nouns like country names and English cognates that could be recognized receptively even by Novice
High/Intermediate Low learners on the ACTFL scale were not removed from the frequency list or sought after in compiling the list of textbook vocabulary. Some ostensibly useful words content lexemes were not in the top 5000 (like bantolo:n, “pants”, 6373; fusta:n, “dress”, 6167; badla, “suit”, 6602; tanu:ra, “skirt”, 6684) but were in call-out boxes within book rather than the primary 5,000-word frequency listing. Another shortcoming emerges in examining the vocabulary type analyzed and the contents of the frequency dictionary. As it was not possible to remove every single spoken Arabic item from the frequency data, projections for vocabulary level coverage may have been distorted by omission of Egyptian spoken Arabic vocabulary items. Still, as the subject of this study is the type and quantity of receptive knowledge for reading, it does not seem likely students at this level or below an advanced level of language study would access or read texts in any variety other than MSA. The corpus used to compile the Dictionary did contain a portion of transcribed spoken data and final counts may have been influenced by the spoken/Standard hybrid form that occurs in the context of online forums; however, but the large majority of texts (newsmedia, news editorials, learned prose, and literature) are still considered to be contexts appropriate mostly for MSA and should have an effect on the corpus as a whole and the final frequency counts within the Dictionary. Even if omitted vocabulary items were added to the total, estimations for vocabulary levels suggested by Van Mol (2000) and Al-Batal (2006) point to the importance of ensuring greater coverage of the 3,000 most frequent words within the Arabic language as a whole. Even as lower levels of text coverage have not been calculated for Arabic, the large gap between the vocabulary knowledge necessary to read a newspaper or novel in English and the size of vocabulary hypothesized to cover the same types of texts
implies this rate may be higher for Arabic than English. Comparing estimations for text coverage by Nation (2006) and Van Mol (2006) for Arabic and English indicate that a minimum projected text coverage rate of 77% may require more than 1,000 word families or 1,800 words as indicated by Nation (2001). The eventual acquisition of the 8,500 word minimum vocabulary knowledge to access newspapers is complicated by the number and type of vocabulary items presented within the textbook series, as well as priorities of teachers, students, and programs involved in using these materials. The following section will address these concerns, and provide pedagogical recommendations to fill perceived gaps.

Pedagogical recommendations

Despite the clear limitations presented by a methodology that may overlook important vocabulary items, analyzing the most prominent textbook vocabulary in light of actual language use patterns allows for a rudimentary gauge of textbook vocabulary’s utility vis-à-vis linguistic reality, as well as providing evidence for the degree of synonymy in the language as a whole (Baker: 1987; Van Mol: 2006). Utilizing corpus-derived frequency data such as the Frequency Dictionary allows TAFL professionals to determine how much of the vocabulary within the textbook will assist students in eventually reaching the minimum level of 8,500 words to access most Arabic newspapers and the daunting task of acquiring upwards of 12,000 discrete lexical items to access literary texts (Van Mol: 2000, 2006). While there are clearly large gaps at the 3,000 and 4,000 word levels, knowledge at these levels should not necessarily be prioritized given projections by Al-Batal (2006) and Van Mol (2000) regarding the importance of the 3,000 or 3,500 word level, even as their projections relating
to proficiency levels differ. A cursory review of literature on text coverage in English and Arabic as well as data regarding an average number of words per family within this study indicates the task of acquiring Arabic as a Foreign Language may be complicated due to the sheer volume of vocabulary knowledge required. Students are faced not only with acquiring and learning divergent orthographic, morphological, and phonological structures, but also with the mammoth task of acquiring a large amount of vocabulary in a language without a great deal of Arabic-English cognates. Due to the vocabulary size necessary for receptive activities, 100-word increments at the 1-999, 1000-1999, and 2000-2999 word levels with the greatest gaps should be prioritized in direct instruction.

Analysis of divergences between corpus-mined data and textbook vocabulary indicates language patterns present in the textbook may not always reflect the most widespread use of a specific lemma or lexeme. The diglossic or multiglossic nature of Arabic complicates this reality, as a certain word or form may not be seen as frequent or salient within the linguistic knowledge of a given Arabic language teaching professional. However, the range and size of text types and precision used to determine the final frequency list within Buckwalter and Parkinson’s Frequency Dictionary provides empirical data on synonyms and forms not found within the textbook but may still be useful for students’ general knowledge as they advance their studies. The broken plurals and divergent forms presented in the final section of this study could be utilized as a guide for actual patterns of language use, but perhaps as a way to point students to the most frequent form if there is a choice between two attested forms. Instructors may explain that the broken plural for “students” is ṭula:b, but the alternate form of ṭalaba may also be encountered as students advance their vocabulary
knowledge and interact with speakers from different regions in the Arabic-speaking world. For forms not attested within the frequency data, alternative and more frequent forms should be investigated and used to augment or substitute for items within the textbook vocabulary not found within the frequency list. These lexemes are less likely to be encountered, and should not be expected to be known productively at the novice level. When time and resources are limited, direct instruction of selected high-frequency items at the 1,000-3,000 word levels not found within the textbook may yield great gains for students in terms of eventual receptive and productive knowledge. This knowledge should be filtered through the lexical contents of Part Two of the Al-Kitaab series if applicable. A future study could also examine the contents of Parts One and Two of the series in isolation and in tandem.

Once students have obtained the appropriate level of vocabulary knowledge, a limited program of extensive reading should be introduced. Extensive reading (ER) is a pedagogical method modified from independent or sustained silent reading programs common for first language learners. According to Day and Bamford (2002), ER involves reading linguistically appropriate and self-selected texts on a wide variety of topics focused on general meaning rather than grammatical form. According to this model, students are generally encouraged to read in the foreign language outside of the classroom setting and as frequently as possible. Texts should be enjoyable and reading becomes “its own reward.” However, this is not to say frequency alone should guide pedagogical priorities. Teachers and TAFL professionals should utilize data culled from the Frequency Dictionary and this study as is appropriate for the linguistic, affective, and instrumental needs of their learners and the Arabic program in
question. At the very least, knowledge of frequency gives curriculum designers and program coordinators options.

**Directions for future research and conclusion**

In 1980, Meara lamented the state of vocabulary research in the field of SLA at large was an “area where the sort of research work out that has been carried out is far from satisfactory, and where a large number of questions still remain to be answered” (p. 19). More recently, Al-Batal (2006) echoed this sense of consternation with the state of vocabulary research in TAFL, as acquisition of vocabulary remains “perhaps the most challenging aspect of learning Arabic” and must be given a “more central part in Arabic curricula, classroom activities, teaching materials, and research projects” (p. 332). Part one of this study represents a small step towards modeling the type and quantity of receptive lexical knowledge a student may possess after completing approximately one to 1.5 years of Arabic using Part One of the *Al-Kitaab* series. Little is known about the Arabic L2 mental lexicon, and data available from Hebrew and general SLA literature points to the role of the root as well as the stem/word as a basic unit in the Arabic L2 mental lexicon. Research regarding the organization and basis of the Arabic FL mental lexicon from the psycholinguistic perspective represents new territory for SLA researchers and professionals hoping to gain insight into the role of the root and word throughout the acquisition process as well as vocabulary acquisition in general. Future research could also remedy the theoretical nature of word families within this study and make projections for maximal and minimal knowledge that could be obtained in light of knowledge of derivational and inflectional processes within the textbook. Testing these projected minimal and maximal word families on foreign learners of Arabic may lead
to a more precise depiction of vocabulary knowledge after completing the textbook series. The issue of receptive and productive vocabulary knowledge remains unexplored in the TAFL context despite advances in the wider field of SLA. Development of a global assessment tool after Nation’s Vocabulary Levels Test (2012) to measure receptive vocabulary knowledge based on frequency data should similarly be given priority. Productive vocabulary knowledge captured by an Oral Proficiency Interview (OPI) may not represent the full extent of a student’s linguistic knowledge. A student may be able to access more vocabulary through reading and listening than can be observed productively in oral and writing production tasks, especially as studies point to a strong relationship between vocabulary knowledge and reading ability (Carver: 2000, 2003 cited in Grabe: 2009; Qian: 1999). The development of such a tool would allow researchers to track vocabulary development over the lifespan of a learner of Arabic as a Foreign Language, including the impact of at-home instruction and the study abroad experience on a given student’s knowledge. While the Frequency Dictionary represents a major stride in cataloging and recording a snapshot of the language as a whole, researchers should work towards cataloguing and increasing the accessibility of larger corpora and encourage the use of these resources for data-based research within the TAFL field. The issues of lexicon and frequency represent a powerful but unaddressed approach within TAFL as a whole. While this is not the only valuable approach to take from the general field of SLA, the application of this approach to TAFL presents both an opportunity and a challenge for curriculum designers, researchers, and TAFL professionals to move beyond intuition and towards an empirically
based approach to the basic issues of curriculum design and instruction within the TAFL field as a whole.
APPENDIX A

MINIMAL WORD FAMILIES FROM AL-KITAAB PART ONE, ALPHABETICAL ORDER

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>اب</td>
<td>Ab</td>
</tr>
<tr>
<td>آباء</td>
<td>Åbae</td>
</tr>
<tr>
<td>أبتدائي 2949</td>
<td>Abtadi</td>
</tr>
<tr>
<td>ابن 1151</td>
<td>Abn</td>
</tr>
<tr>
<td>ابناء</td>
<td>Abnæa</td>
</tr>
<tr>
<td>ابن عمّ</td>
<td>Abn amm</td>
</tr>
<tr>
<td>أسبوع 393</td>
<td>Abw</td>
</tr>
<tr>
<td>أسابيع 393</td>
<td>Abw</td>
</tr>
<tr>
<td>اسم 89</td>
<td>Asm</td>
</tr>
<tr>
<td>اسماء 89</td>
<td>Asmæ</td>
</tr>
<tr>
<td>اجر</td>
<td>Äjżr</td>
</tr>
<tr>
<td>استأجر</td>
<td>Ästæjżr</td>
</tr>
<tr>
<td>يستأجر</td>
<td>Æstæjżr</td>
</tr>
<tr>
<td>الاستئجار</td>
<td>Al-استئجار</td>
</tr>
</tbody>
</table>

Note: Only lemmas attested in the *Frequency Dictionary* among the 5,000 most common lemmas in the language are annotated with a corresponding frequency ranking. Words grouped in clusters represent disparate hypothetical word families. Words thought to be learned and accessed through the consonantal root are listed as such while lexical items that may not be easily analyzed by novice students according to the traditional root approach are listed as full lexical items without recourse to the root. These items are listed alphabetically.
احتجاج 538
أحياناً 358
أخ 66
أخوة 66
أخ ر
تأخر 2221
تأخرت (أنا)
متأخر ون، بن 2016
آخر
آخر 38
آخرى
أخرون
آخر + اسم مفرد 474
أخيراً
أخ ذ
أخذ 198
أخذت (المبني المجهول)
الأخذ 1785
أخلاق 1807
ادب 1195
إدارة الأعمال
ذا + الماضي 43
أراد 126
أريد
ارز 4015
الاقتصاد 969
استاذ 562
استاذة 562
استاذات
أسرة 664
أسر 664
أسرته
اعدادي
أغنية 838
أغاني
أغاني

ال alm 1

اليأكل 1338

التي 18

الآن 125

الأمم 53

ألف 111

الألف 111

أم 163

أمهات 163

أما... ف 134

 أمس 237

 امرأة 1052

الام المتحدة

انتداب

انسان 204

 إن ف
استتأنف

يستتأنف

الاستتأناف 2852

أو 23

أوّل 41
أولى 41

ايّ 46؟

أيضاً 146

ب

باب 298
 أبواب 298

بالإضافة 382

117 بسبب

بجانب 208

بفضل 705

بكم؟

بالنسبة

بحاجة الي
507 بحر
507 بحار

بخير

بد أ 172 بدأ
بد أ بدأ
البدء 1023

ب د ل
تبادل
تبادل
التبادل 1602

بدل
بدل

بارد 1380
بريد 1812

برنامج 161
برنامج 161

برنطة
برانط

سرعة

بطن 2080
بعد الظاهر

بعض + اسم جمع في أضافة 50

ب ق ي
بقي 257
يبقي
البقاء 1555
بقية 928

 بكالوريوس

بلوزة
بلوزات

ب لوفر
ب لوفرات

بننت 252
بنات 252

بنطلون
بنطلونات

بنوك 625
بنوك 625

بن ي
بنية 4578
بنيات
ب ن ن
بني 4164
نية

يت 104
بيوت 104

بيجاما
بيجامات

ب ي ض
أبيض 564
بيضاء 564
بيض 564

ت

تاريخ 286
تجارة 886

تجربة 587
تجارب 587

ت ر ج م
ترجمة 2228
ترجمات
مترجم
مترجمون/ين
ترجمة
مترجمات

التزاج
ت رك
ترك 472
بترك
الترك

تصرف 1657
تصرفات 1657

التصوير 1880

تلك 129

ت م م
تمّ 78
بتمّ

تنورة

تي شرت
تي شرتات

ث

ثلث 1504
ثلث 2359

ثلج 3500

ثمّ 108
ثقة
الثاني 1347
وثاني 1455

الثانية العامة 741

الثقافة 519

ج
جار 1550
جيران 1550
جار
جارات

حاكيم
حاكيمات

جسد 1226
جسد
ابن
جسد
جذب

171

الجذر 2610
جريدة 1339
جرائد 1339
جزء 524
أجزاء 5241

جزيرة 1733
جزر 1733

جلابة
جلالبيب

جلالة الملك

جلس 924
جلس
الجلس 3053

جامعة 210
جامعات
جمع 1107
الجامعة 955
جامعة 2812
جامع
جميعاً 82
مجموعة 293
مجموعات 293
اجتماع 1974
جتماع
الاجتماع 475

جملة 1422

جنوب 416

الجو 667

جوائز 2995
جوائز السفر 2995

جي أ
 جاء (جنت) 109
 يجيء
المجيء

جيب 2512
جيبات

جيش 337
جيوش 337

جيداً 488

ي

الحج 2463

حج ب ب
 أجب ا 495
 الحبّ 209

حج د د
 حادث 1431
 حوادث 1431
 حدث 408
 يحدث 453
 يحدث الحديث 369

حديث أحداث

حج د د
 حدث 4115
 يحدث 684

حديثة 1624
 حدائق 1624

حذاء 3637
 أحداث

حرارة 1370
 حرام
الحرية

ح س ن
أحسن 776

ح ص ل
حصل 445
حصول 846
حصلت

ح ض ر
حضور 809
بحضر
الحضور 700

ح ف ظ
أحفظ 1228

ح ق ق ق
الحقيقة 322
الحقوق 84
حقوقي 537
حقوق 707

ح ك م
حكم 1694
بحكم
الحكم
حكومة 103
حكومات 103
حلم 637
أحلام 637

حمام 2309
حمامات

ح م ر
أحمر 927
حمراء 927
حمر 927

ح م ل
حمل 334
يحمل
الحمل 1256
موا لي 808

الحياة 94
حيث 79

خ
خاص 135
خالة 2926
خالات
خال 3188
أخوات
خبر 285
أخبار 285
خبرة 1125
الخجل 3137

خ ر ج
خرجت (أنا)
الخرج 4180
خرج 345
خرج 872

الخريف 3034

خ ض ر
خضار 3965
أخضر 987
خضراء 987
خضر 987

خطب ل

خ ل ف
اختلاف 1109
يختلف
الاختلاف 1682
مختلف 259
مختلفون/ين
خليفة 2751
خلفاء
الخميس 1115
اختار 1205
يختار الاختيار 959
د
دائماً 283
دجاج 2665
دخ ل
دخلت
أدخل 297
الدخول 595
دخ ن
دخن
دخَن
الدخنين 2333
درجة 473
درس
درس
درس 1066
دروس 1066
درّس
يدرس
مدرسة 303
مدارس 303
داراسي 1718
دراسة 287

دفتر 3904
دفتر

دقيقة 424
دقائق 424

دكتور 253
دكتورة 253
دكتوراه 4106

الدم 363

الدنيا 494
الدين 333
ذ

ذاكر
ذكراً (أنا)
المذكارة
ذكر (أنا) 1217
ذكر 266
ذكر
الذكر 1002
ذكي 2496
أذكياء
ذلك 36
ذهب
ذهب (أنا) 489
ذهب 1692
ذهبية ر راحة 1476
الرأس 215
رئيس 47
رؤساء 47
رأي 227
آراء 227
رأي أرى 75
الرب 194
ربنا
رب ع
رابع 714
الربع 1291
ربع 1319

الربع 1943

رتب
رتب
يرتب
الترتيب 1534
ترتيبات

رجل 92
رجال 92

الرجل 4356

رجو 964
رجاء 2388

رحب 2022
رحب
الرحيب 2889
<table>
<thead>
<tr>
<th>الرقم</th>
<th>الرخص</th>
<th>رسائل</th>
<th>رسائل</th>
<th>أرسال</th>
<th>أرسال</th>
<th>الرسوم</th>
<th>الرسوم</th>
</tr>
</thead>
<tbody>
<tr>
<td>2640</td>
<td>108</td>
<td>2640</td>
<td>433</td>
<td>433</td>
<td>1592</td>
<td>1569</td>
<td>1569</td>
</tr>
<tr>
<td>1748</td>
<td>108</td>
<td>1387</td>
<td>1387</td>
<td>1387</td>
<td>1387</td>
<td>1387</td>
<td>1387</td>
</tr>
<tr>
<td>4794</td>
<td>935</td>
<td>935</td>
<td>935</td>
<td>935</td>
<td>935</td>
<td>935</td>
<td>935</td>
</tr>
<tr>
<td>3077</td>
<td>708</td>
<td>708</td>
<td>708</td>
<td>708</td>
<td>708</td>
<td>708</td>
<td>708</td>
</tr>
<tr>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
</tr>
<tr>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
<td>1182</td>
</tr>
</tbody>
</table>
الرياضية 704
ز
ز ه م
الازدحام
ز ر ق
أزرق 1754
زرقاء 1754
زرق
زميل 1428
زلاء 1428
زميلة
زميلات
زهري
زهرة
زوج 464
أزواج 464
زوجة 496
زوجات 496
زور
زار 1563
بزور
الزيارة 490
زيت 1745
س

س149

ساعة185
ساعات185

السباحة2920

سادس1237

السبت1148

السحور

سرير1951

س ع د

الممساعدة692

أساعد1056

س ف ر

السفر1027

تسافر1672

سفر1015

سفراء1015

س ك ن

أسكن1328

السكن2342

سلام188
سلطة

س ل ك
سلوك 1727

س م ح
سمح 787
يسمح
السماح 2320

س م ر
أسمر 4452
سمراء 4452
سمر 4452

س م ع
يسمع 2143
الاستماع 3571

سمك 1916

السماء 728

سنة 69
سنوات

س ه ر
سهر
يُسُهِر
السهر 4501

س و د
<table>
<thead>
<tr>
<th>طريقة</th>
<th>سرعة</th>
</tr>
</thead>
<tbody>
<tr>
<td>أسود 682</td>
<td>سوداء 682</td>
</tr>
<tr>
<td>سود 682</td>
<td></td>
</tr>
<tr>
<td>سورة 2238</td>
<td>سورة 2238</td>
</tr>
<tr>
<td>سوف 460</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>قطعة</th>
<th>عدد</th>
</tr>
</thead>
<tbody>
<tr>
<td>السياحة 1566</td>
<td></td>
</tr>
<tr>
<td>سيارة 251</td>
<td>سيارات 251</td>
</tr>
<tr>
<td>سيّدة 189</td>
<td>سادة 189</td>
</tr>
<tr>
<td>سيدة 968</td>
<td>سيدة 968</td>
</tr>
<tr>
<td>السينما 1730</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>شاب</th>
<th>عدد</th>
</tr>
</thead>
<tbody>
<tr>
<td>576</td>
<td>شاب</td>
</tr>
<tr>
<td>264</td>
<td>شباب</td>
</tr>
<tr>
<td>576</td>
<td>شبان</td>
</tr>
<tr>
<td>467</td>
<td>شارع</td>
</tr>
<tr>
<td>467</td>
<td>شوارع</td>
</tr>
<tr>
<td>4148</td>
<td>شبايك</td>
</tr>
</tbody>
</table>
الشتاء 1722

ش ج ع
شجع 3032
يشجع
الشجع 1922

شخص 241
أشخاص 241

الشطرنج

ش ر ب
بشرب 2081
شوربة

ش ر ف
أشرف 2903
يشرف
الاشراف 2280

شرق 386

الشرق الأوسط

شركة 118
شركات 118

ش ر ك
اشترك 2874
يشترك
الاشتراك 2970
ش ر ي
اشتُر 1265
يشتري
الشراء 1210

ش ع ر
أَشْعَر 508
الشعر 1111

الشعر 947

ش غ ل
الشغُل 749
مشغول 2494
مشغولون/ين
مشغولة
مشغولات
انشغال
ينشغال
الانشغال
شغُل 2104
يشغُل

ش ق ر
أَشْقُر
شرقاء
شقر

شقة 2321
شقق

شكل 132
أشكال 132
شمال 624

ش ه د
يشاهد 1789 المشاهدة 2090

ش ه ر
شهر 158
 أشهر 158
شهر 158

شورت
شورتات

شيء 39
أشياء 39

شي ب
أشيب
شيئا
شيب

ص

ص ب ح
أصبح 178
يصبح

صاحب 190
صف 2099
صف 905
صفوف 905
صف
صفات
الصلاة 635
ص ل ع
أصلع
صلع
صورة 120
صور 120
ص و م
صام
يصوم
الصوم 4223
الصيام 3470
صيدلية
صيدليات
الصيف 1121
ض ض
ضابط 1008
ضباط 1008
ضابطة
ضابطات
ضروري 1214
ط
طاولة 1857
طلالات
طائرة 973
طائرات 973
طابق 3892
طوابق
طالب 270
طلاب 270
طالبة
طالبات
الطبيب 1529
ط ب خ
مطبخ 2999
مطابخ
طبخ
طبخ
طبخ
الطبيخ 4669
طبخاً
طبعاً 340
طفل 174
طفلة 174
أطفال 174
الطفولة 1949

طقس 2657
طلب 457
يطلب
الطلب 575

طوال 1343

ط و ع
استطاع أن 235
يستطيع أن

ظ
ظ ن ن
ظَنَّ 1038
ظننت
يطن
الظنّ 2546

ع
عاصمة 696
عواصم 696

عاطفي 2897
عالية 505
عام 62
أعوام 62

عائلة 850
عائلات 850

ع بر
اعتيار 341
يعتبر
الاعتبار 758

ع د
الأعداد 119
عدد 503
عدد 119
عدد 2765
عدد 113

استعداد 2994
يستعد
الاستعداد 1177

عربي 45
عرب 45
الأعراب

عرف 49
المعرفة 610
تعرَّف
يتعرَّف
النَّتَّعَرَف 1966

عَشاء 2239

عطلة 2947

عطل

علاقة 219

علاقات 219

عَلِم م

علم 548
علم 377

يَعْلَم

العلم 515
علم 3450

يَعْلَم

التعليم 388

علم الاجتماع

علوم السياسية

علم النفس

العالم 93

عَمَّ= عن مَاذا

عَمَّ= عمَّ 788

أعمَّ
عمر 281

ع م ل
عمل
يعمل 137
العمل 48
أعمال 48

عام 88
عامة

العامة

ع م ي
أعمى 2758
عمياء 2758
عمي

عندما + فعل 166

ع ن ي
يعني 751
معنى 436
المعاني 436
معانٍ 436

ع و د
عاد 123
يعود العودة 404
معيد معيدة
عادة 734
اعتماد 3527
يعتاد
عيد 647
أعياد 647
ع ي ش تعيش 326
عين 130
عيون 130
عينت 3579
غ
غائم
الغداء 2774
غدا 775
غرب 695
الغيرة 2581
غرفة المعيشة

غض ب غضب 4154 يغضب الغضب 1165

غالي 785 غالي 785

غي ب تغيب يتغيب التغيب غاب 1607 يغيب العياب 1212

غي ر غير 1648 يغير التغيير 592 تغيير 952

فر 626 افراد 626
فرصة 410
فرص 410

فر ق ر تفر ق
تفر ق
يتفر ق

الفشل 1452
يفشل
الفشل 3168

القصصي

فصل 1856
فصل 1856

فصل 1915
يفصل
الفصل 1377

فصلي
فصليّة

فتر
يفتر
يفتر
الفطور
الافظرار 4335

فعل ل
فعل 254
أفعال 254
الفاعل 1926
المفعول به
فعل.

فقط 274

فك ر
فك 622
فكرة 471
أفكار 522

فلس 2706
فلوس 2706

فلسطينية 107
فلسطينيات

فندق 1352
فنادق 1352

فهم 580
يفهم
فهم
الفهم 1388
فواكه
فوطة
فوطة

فوق 348
في 3
في أيّة؟
في أيّة ساعة
في رايه
ق

القادم 500
القادمة

ق ب ل
القبول 1300

قابل 4254
يقابل
مقابلة 1818
مقابلات 1818

قبل 68

استقبل 1726
يستقبل
الاستقبال 1583
القدس
قديم 499

ق ر أ قرآً 727
القراءة 923

 قريب 476
أقارب 1354

 ق ر ر قرر 767
يقرر 211
قرارات 211

استقرر 2527
يستقرر
الاستقرار 997

قسم 683
قصد 1337
يقصد
القصد 1330

قصة 414
قصص 414
قضى

القضاء

انقطع

الانقطاع

القلب

قليلاً

قليل

قميص

قمصات

مقهى

مقاهي

المقاهي

القواعد

آقام

أقمت

الإقامة

قام
يقوم القيام 651

لك

لك ب ر
أكبر 195

لك ت ب
كتب 357
كتاب 196
مكتبة 565
مكتاب 565
مكتبة 1830
مكتبات 1830

يكتب
الكتابة 966
كاتب 719
كتاب 719

لك ث ر
كثر أ 55
كثر 64

كرافات
كرافات
كراء السلة
كراء القدم
كرة الطائرة

كل 19

كم 336

كما 52

كنِّية 2056

كِّنَّيسَة 2056

و 10

كانكنتكانت

لَّكَوْن 2630

ينَكَوْن

مكان 179

اماكن 179
كيف؟ 67

ل (in order to)

ل (possessive)

لآن + جملة اسمية 57

لب س

للس 2617

ليللس

اللبس 4312

لبناني 212

لبنانيون/ين

لحم 1518

لح ق

التحق ب (أنا)

لذلك 465

لسان 1026

لع ب

لعب 619

لعب

اللعب 843
لمة؟

ل م ح
لمح
بلمح
التميحي

لون 498
الوان 498

لي س
ليس 59
ليس هناك
ليس له
ليس لها
ليس فيه
ليست

لي ل
ليلة 518
ليالي
ليالي 518
للليل

؟

231
محافظة 134
الماضي 203
بتدأ
مبارك 1535

مباراة 419
مبارات 419

محدود 221
محدود

متخصص 1519
 متخصص/ين
 متخصصة
 متخصصات

منفوق 3448
منفوق/ين

م ت ع
استمتع
يتمتع
الاستمتع
ممنع
المساء 642

مرور 528

مرات 74

مرُة 74

مرحومة 3576

مرحوم 379

مدينة 144

مدٌّكر

محافظة 660

محافظات 660

مدير 1132

مجلة 1132

مجلات 1132

المثنى 1631

مثل

المساء 642
مستقبل

المسرح

مسلسلات1704

مشكلة181

مشاكل 181

ممس

مشهورين2563

مشهورون

مصرية381

مصرية بين 381

المضارع

المضارع المرفوع

المضارع المنصرب

مطر

مطر

مطر
مع 17
مع من؟
معظم
مفرد
4028
مقالة
مقالات
1022
1022
ن
يمكن
يمكنني أن
يمكنك أن
يمكنه أن
ملابس داخلية
ملك
ملك
ملك
ملك
ملك
ملك
ملك
ملك
ملك
ملك
من أين؟
من اصل
من اللازم

المشاكلة 1316

منتصف 1649

منحة 3537

منحات

منذ 175
منذ سنتين

منطقة 83

مناطق 83

م ن ي
تمنى 451
يثمنى

مهم 249
مهمون/ين

المهم هو
الموسيقى 1440

مؤسسة 290

موضوع 95

مواضع 95

مؤطّف 915

مؤظفة

مواعيد 1005

مؤنث

ناس 114

نبيّ 813

ن ج ح

نجحت في 954

أنجح

النجاح 617

النادي 411
النوادي

ن ز ل
نزل 673
ينزل
النزل 8

نوزل

نساء 511

النسبة

ن ش أ
نشأ 2634
نشأ

نشوءة

نصف 607

نصيحة 1866
نصائح 1866

نظارات

نفس ال

ن ق ل
نتقل 1512
نتقل

الانتقال 2180

النهار 980

ن و
انتهى
ينتهي
الانتهاء

ن و م
نام
1268
نمت
ينام
النوم
930

ن و ي
نوى
3128
ينوي
النية
1712

لي
ه
هجر
4922
يهجر
الهجرة
1760

هدية
1438
هدايا
1438

هل؟
72

هنا
159
هناك
77
ه ن د س
الهندسة 1134
مهندسون
مهندسين

هواية
هوايات

و
وجد
وجد 139
يجد
أجد
ما وجدت
(انا)

وحيد
وحيدة
الوحدة 247
الأحد، لا أحد

ورقة 558
أوراق

وزارة 217

الوزن

وص ل
وصل 183
يصل
الوصول 638
تتصل
اتصل 722
يتصل
الاتصال 493

وطن 373
أوطن 373

وظيفة 1157
وظائف 1157

وفاة 1544

وقت 98
أوقات 98

وقت توقف 936
يتوقف
التوقيف 2541

ولد
والد 559
والدي
والدة 559
والدتي

ولد 362
أولاد 362

ولدت
و ل ي تو لى 2014
بتولى النّولى

ولاية 276
ولايات 276

الولايات المتحدة الأمريكية

ي

اليد 148

ي ق ظ

يوقظ

يوم 26
أيام 26
اليوم 26
## APPENDIX B
TEXTBOOK VOCABULARY ARRANGED BY FREQUENCY

### 1-1000

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>ال 1</td>
<td>1</td>
</tr>
<tr>
<td>في 3</td>
<td>3</td>
</tr>
<tr>
<td>من 4</td>
<td>4</td>
</tr>
<tr>
<td>ل 5 (possessive)</td>
<td>5</td>
</tr>
<tr>
<td>كان 10</td>
<td>10</td>
</tr>
<tr>
<td>مع 17</td>
<td>17</td>
</tr>
<tr>
<td>التي 18</td>
<td>18</td>
</tr>
<tr>
<td>كل 19</td>
<td>19</td>
</tr>
<tr>
<td>أو 23</td>
<td>23</td>
</tr>
<tr>
<td>يوم أيام اليوم 26</td>
<td>26</td>
</tr>
<tr>
<td>لم 27</td>
<td>27</td>
</tr>
<tr>
<td>بعد 34</td>
<td>34</td>
</tr>
<tr>
<td>ذلك 36</td>
<td>36</td>
</tr>
<tr>
<td>آخر اخري 38</td>
<td>38</td>
</tr>
<tr>
<td>شيء اشياء 39</td>
<td>39</td>
</tr>
<tr>
<td>أول أولى 41</td>
<td>41</td>
</tr>
<tr>
<td>إذا + الماضي 43</td>
<td>43</td>
</tr>
<tr>
<td>عربي عرب 45</td>
<td>45</td>
</tr>
<tr>
<td>أي؟ 46</td>
<td>46</td>
</tr>
<tr>
<td>رئيس رؤساء 47</td>
<td>47</td>
</tr>
<tr>
<td>العمل أعمال 48</td>
<td>48</td>
</tr>
<tr>
<td>بعض 50</td>
<td>50</td>
</tr>
<tr>
<td>كما 52</td>
<td>52</td>
</tr>
<tr>
<td>الأ 53</td>
<td>53</td>
</tr>
<tr>
<td>كثيراً 55</td>
<td>55</td>
</tr>
<tr>
<td>لأن 57</td>
<td>57</td>
</tr>
<tr>
<td>ليس 59</td>
<td>59</td>
</tr>
<tr>
<td>عام أعوام 62</td>
<td>62</td>
</tr>
</tbody>
</table>

أكثر ٦٤
أخ ٦٦
كيف؟ ٦٧
قبل ٦٨
هل؟ ٧٢
مرة مرات٧٤
هناك٧٧
ثم٧٨
حيث٧٩
جميعاً٨٢
منطقة مناطق٨٣
الحقوق٨٤
عام٨٨
اسم اسماء٨٩
رجل رجال٩٢
العالم٩٣
الحياة٩٤
موضوع مواضيع٩٥
وقت أوقات٩٨
قام١٠٢
حكومة حكومات١٠٣
بيت بيوت١٠٤
ثم١٠٨
جاء١٠٩
قلب١١٠
ألف الألف١١١
الاعداد١١٣
ناس١١٤
بسبب١١٧
شركة شركات١١٨
عدد الاعداد١١٩
صور صور١٢٠
عاد١٢٣
تلك 129
الآن 125
أراد 126
عين عيون 130
شكل أشكال 132
أما ف 134
خاص 135
وجد 139
بخير 140
مدينة مدن 144
أيضا 146
اليد 148
س 149
شهر أشهر شهور 158
هنا 159
برنامج برامج 161
أم أمهات 163
عندما 166
 جدا 171
بدأ 172
كلمة كلمات 173
طفل طفلة أطفال 174
منذ 175
أصبح 178
مكان أماكن 179
مشكلة مشاكل 181
وصل 183
ساعة ساعات 185
سلام 188
سيد سادة 189
صاحب أصحاب 190
الرب 194
أكبر 195
خبر أخبار 285
التاريخ 286
الدراسة 287
مؤسسة 290
مجموعة مجموعات 293
باب أبواب 298
مدرسة مدارس 303
الحقيقة 322
الدين 333
حمل 334
كم؟ 336
جيش جيوش 337
طبعاً 340
اعتبر 341
خرج 345
فوق 348
جري 353
كتب 357
أحبانا 358
ولد أولاد 362
الدم 363
صحيح 364
الحديث 369
وطن أوطان 373
لمذا؟ 375
قليلًا 376
علم 377
مدير 379
مصري مصريون 381
بالإضافة إلى 382
شرق 386
أسبوع أسابيع 393
صديق أصدقاء 398
حدث 408
فرصة فرص 410
 قصة قصص 414
جنوب 416
مباراة مبارات 419
دقيقة دقائق 424
ملك ملوك 425
رسالة رسائل 433
معنى المعاني معان 436
لغة لغات 441
ل (in order to) 442
حصل 445
 أقل 446
تمني 451
تحدث 453
طلب 457
الصحة 458
سوف 460
زوج أزواج 464
شارع شوارع 467
ترك 472
درجة 473
آخر 474
قريب 476
صحيفة صحف 486
جداً 488
الزيارة 490
الاتصال 493
الدنيا 494
زوجة زوجات 496
لون اللوان 498
قديم 499
القادم 500
ب)حاجة 621
شمال 624
بنك بنك 625
فرد افراد 626
القواعد 630
الصلاة 635
حلم أحلام 637
الوصول 638
المساء 642
عيد أعياد 647
القيام 651
محافظة محافظات 660
أسرة أسر 664
الجو 667
نزل 673
أسود سوداء سود 682
قسم 683
التحديد 684
المستعمرة 692
غرب 695
عاصمة عواصم 696
الحضور 700
الرياضة 704
بفضل 705
حقًّا 707
رفض 708
رابع 714
كاتب كتاب 719
أجنبي أجنب 723
السماء 728
عادةً 734
الشغل 749
يعني 751
سيدة سيدات
اقتصاد
طائرة طائرات
النهار
أخضر خضراء خضر
الاستقرار

1000-1999

الذكر
موعد
ضابط ضبط
سفير سفرا
مقالة مقالات
البدء
السفر
ظل
امرأة

دروس دروس
جمع
اختلف
الخميس
الصيف
خبرة

مجلة مجلات
مهندس
قضي
السبت
ابن

وظيفة وظائف
الغضب
الاستعداد
رواية روايات
ثانويّ1455
راحة1476
الثلاثاء1504
انقل1512
لحم1518
متخصص 1519

الطب1529
الترتيب1534
مبروك1535
وفاة 1544
جار جبران 1550

البقاء1555
زار1563
الرسم1569
الاستقبال 1583
أرسل 1592
التبادل1602
غاب 1607

حديقة حدائق 1624
مثل1631
تعزف1634
غير 1648
منتصف 1649

تصرف تصرفات 1657
الاختلاف 1682
ذهبى1692
حكم 1694

مسلسل مسلسلات 1704

النهاية 1712
دراسي 1718
الشتاء 1722
مستقبل 1726
سلوك 1727
السينما 1730
جزيرة جزر 1733
زيت 1745
الرسال 1748
أزرق رفقاء 1754
الهجرة 1760
الأخاذ 1785
أخلاق 1807
بريد 1812
 مقابلة مقابلات 1818
مطعم مطاعم 1823
مكتبة مكتبات 1830
رغب 1838
فصل فصول 1856
طولة 1857
نصيحة نصائح 1866
التصوير 1880
فصل 1915
سمك 1916
التشجيع 1922
الفاعل 1926
الريغ 1943
الطفولة 1949
سرير 1951
محاضرة محاضرات 1961
التعريف 1966
اجتماع 1974
2000-2999
تولي 2014
متاخر 2016
رحّب 2022
كنيسة كناتس 2056
البطن 2080
المشاهدة 2090
أصفر صفراء صفر 2099
شعل 2104
إجازة إجازات 2114
انقطع 2170
الانتقال 2180
صادقة 2206
مقهى مقاهي 2218
تأخر 2221
ترجمة 2228
سورة سورة 2238
عشاء 2239
السماح 2320
الشرف 2280
حمام 2309
التدخين 2333
ثالث 2359
الرجاء 2388
الهندسة 2448
الحاج 2463
مشغول 2494
ذكى 2496
استقر 2527
التوقف 2541
اللوه 2546
مشهور 2563
الغربة 2581
الجزر 2610
تكون 2630
نشأ 2634
رخيص 2640
الاستئناف 2852
اشترك 2874
الترحيب 2889
عاطفي 2897
أشرف 2903
سباحة 2920
خالة 2926
بطلة 2947
أيداً 2949
الاشتراك 2970
استعد 2994
جواز سفر 2995
مطيخ 2999
3000-3999
شجع 3032
الخريف 3034
الجلس 3053
النزول 3068
القص 3077
نوى 3128
الخجل 3137
الفض 3168
خال 3188
طقس 2657
دجاج 2665
فلس فلوس 2706
خليفة 2751
أعمى عمى 2758
أعد 2765
الغداء 2774
جامع 2812
الاستئناف 2852
اشترك 2874
الترحيب 2889
عاطفي 2897
أشرف 2903
سباحة 2920
خالة 2926
بطلة 2947
أيداً 2949
الاشتراك 2970
استعد 2994
جواز سفر 2995
مطيخ 2999
3000-3999
شجع 3032
الخريف 3034
الجلس 3053
النزول 3068
القص 3077
نوى 3128
الخجل 3137
الفض 3168
خال 3188
الرجوع 3412
الانقطاع 3439
منفوق 3448
الصيام 3470
الملك 3477
ثلج 3500
اعتداء 3527
منحة 3537
الاستماع 3571
مرحوم 3576
حذاء 3637
خضار 3965
4000-5000
ارز 4015
مفرد 4028
دكتوراه 4106
حدّد 4115
ماجستير 4137
الثانوية (العامة) 4147
غضب 4154
بني 4164
التخرج 4180
الصوم 4223
الافطار 4335
قابل 4254
الرجل 4356
أسرم سمرة سمر 4452
السهر 4501
بناءة 4578
قميص 4630
الطيب 4669
فقط

4791

هاجر 4922
REFERENCES


