# Arabic Word Learning in Novice L1 English Speakers: Multi-modal Approaches and the Impact of Letter Training in the Target Language Script

**MEREDITH H. CICERCHIA, MSc.**

Thesis submitted to the University of Nottingham

for the degree of Master of Philosophy

DECEMBER 2022

**Abstract**

This thesis explores early Arabic word learning by beginner level native English speakers who have no prior exposure to the language. While Arabic is considered a difficult language for English speakers to learn, very few studies focus on Arabic as a Foreign Language (AFL) vocabulary acquisition in beginners, despite vocabulary's central role in language learning. The present research encompasses two separate word learning studies which employed multi-modal learning tasks in a language lab setting to show that novice adult learners can acquire Arabic vocabulary given minimal exposure to target language input accompanied by images and audio. In the first study, response time and accuracy data were used to explain performance on a word learning task and probe difficulty drivers in the target language word set. Findings suggest the number of letters and syllables a word contains can explain response time while the number of Arabic-only phonemes it contains can have a significant impact on accuracy. The second study provided a subset of participants with an Arabic letter-training session and utilized written word forms in a modified version of the target language script. Results showed a significant advantage for the letter-training group across all measures of learning. Findings support the use of letter training at introductory level and suggest novice learners can make use of the Arabic script to support form-meaning mapping in early vocabulary study. Results complement existing work on multi-modal learning paradigms and are discussed in the context of research for AFL. They may be used to support future study design and inform stimulus selection for vocabulary researchers choosing to work with Arabic, and generally serve to advance our understanding of effective approaches to Teaching Arabic as a Foreign Language (TAFL) to English native speakers.

## Acknowledgements

I want to thank my supervisors Dr. Kathy Conklin and Dr. Dominic Thompson for their guidance and support throughout this project. Without their belief in my work and patience during the end stages, it may never have been completed. I also want to thank Joanna Pullen from the School of English for her help in navigating extension paperwork and the challenges of working through lockdowns, and Ira Karakatsani from the University Counselling Service, for her support and encouragement. I am indebted to the postgraduate students and faculty in charge of the School of English's Psycholinguistics lab, especially Ling Shuping, for their help with data collection and running experiments, and to the University of Nottingham's School of English Research Committee and the Qatar Foundation International for sponsoring my studies. I also want to extend special thanks to Dr. Giselle Khoury, Tim Buckwalter, Gunna Funder Hansen, Dr. Jackie Masterson, Dr. Aman Attieh, and Dr. Ludovic Ferrand for their support and help accessing documents, databases and research to inform my project, and Mahmoud S. Moamenah for his help with recordings. Special thanks to the Psycholinguistics Research Group for providing feedback and a forum for practicing conference presentations. Many thanks to the organizers of the British Council's Arabic Language and Culture Programme for inviting me to participate and present my work at their meeting for Arabic teachers in the UK, and to the British Association for Applied Linguistics for including my research in their research reports newsletter. Lastly, I want to thank my family and friends for their continued support, especially my parents, Sandra and Freddy, for believing in me and always encouraging me to challenge myself and never stop learning.

# Table of Contents

List of Tables………… 7
List of Figures………… 8
List of Appendices………… 9

**Chapter 1: Introduction**………… 10
Organization of the Thesis………… 14
**Chapter 2: The Arabic Language and Word Learning Studies** ………… 18
**The Arabic language**………… 18
Diglossia or Multiglossia? ………… 20
Literary Arabic………… 20
Spoken Arabic………… 21
Reporting Dialects in the Research………… 22
Talking about Arabic – Conventions and Abbreviations………… 24
Are All Native Speakers of Arabic Bilingual? ………… 25
**Arabic as a "Difficult" Language**………… 26
Morphology………… 27
Script and Orthography………… 27
Diglossia………… 33
Cognates and Phonology………… 34
**Resources for Researchers**………… 38
Arabic Corpora………… 38
Frequency Dictionaries………… 42
Text-book Analysis and Assessment Tools ………… 44
Normative Image Databases for Arabic………… 44
**Chapter 3: Arabic Word Learning Research** ………… 48
Additional Challenges………… 48
Modality of Input………… 48
Transliteration and Transcription………… 49
Word Learning Studies………… 53
Summary of Research Reviewed………… 58
**Chapter 4: Multi-modal Arabic Word Learning in Novice L1 English Speakers**………… 60
Abstract………… 60
Introduction………… 61
**Models of Working Memory and Theories of Processing**………… 62
Using Images in Word Learning………… 63
**Multi-modal Learning**………… 65
Incidental Learning Studies………… 65
Deliberate Learning Studies………… 67
Multi-modal in a Classroom Context………… 69
Multi-modal AFL Studies………… 71
Summary of Research Reviewed………… 74
**Present Study and Research Aims**………… 75
Method………… 75
Participants………… 76

Design………………………………………………………………………… 77
Materials……………………………………………………………………… 78
Procedure……………………………………………………………………… 81
**Results**………………………………………………………………………… 83
Preliminary Analyses………………………………………………………… 83
Accuracy………………………………………………………………………… 85
Response Times……………………………………………………………… 86
Results Across Blocks………………………………………………………… 87
Number of Letters, Syllables, and Arabic-only Phonemes………………… 88
Additional Analyses…………………………………………………………… 91
**Discussion**…………………………………………………………………… 91
**Chapter 5: The Impact of Letter Training on Multi-modal Word Learning in Arabic**………………………………………………………………………… 97
Abstract………………………………………………………………………… 97
Introduction…………………………………………………………………… 98
**Orthographic Input and Phonological Awareness**………………………… 100
**Letter Training Studies**……………………………………………………… 104
**Present Study and Research Aims**………………………………………… 107
Method………………………………………………………………………… 108
Participants…………………………………………………………………… 108
Materials……………………………………………………………………… 110
Letters………………………………………………………………………… 110
Words………………………………………………………………………… 111
Images & Audio……………………………………………………………… 112
Letter Training………………………………………………………………… 114
Word Learning………………………………………………………………… 115
Word Recognition Task ……………………………………………………… 116
Recall Task…………………………………………………………………… 116
Delayed Word Recognition and Recall ……………………………………… 117
Procedure……………………………………………………………………… 117
**Results**………………………………………………………………………… 118
Letter Training………………………………………………………………… 119
Immediate Word Recognition -- Accuracy…………………………………… 121
Immediate Word Recognition – Response Times…………………………… 123
Immediate Recall -- Accuracy………………………………………………… 125
Delayed Word Recognition -- Accuracy……………………………………… 126
Delayed Word Recognition – Response Times……………………………… 126
Delayed Recall -- Accuracy…………………………………………………… 126
Word Type (i.e., Number of Taught Letters) – Immediate Word Recognition Accuracy ………………………………………………………………………… 128
Word Type (i.e., Number of Taught Letters) – Immediate Word Recognition Response Times………………………………………………………………… 132
Word type (i.e. number of taught letters) - Immediate Recall Accuracy………… 133
Word type (i.e. number of taught letters) - Delayed Word Recognition Accuracy.. 133
Word type (i.e. number of taught letters) - Delayed Word Recognition Response Times………………………………………………………………………… 134

Word type (i.e. number of taught letters) - Delayed Recall Accuracy……………. 135
Additional Analyses……………..……………………………………………… 135
**Discussion**………………………………………………………………………. 136
Overview…………………………………………………………………..…. 136
Letter-training Results ………………………………………………………. 137
Performance of the Letter-training Low Group…………………………………. 138
Main Study Results………………………………………………………… 140
Performance of the Letter-Training High Versus No Letter Training Group……. 140
Performance on Delayed Measures………………………………………………. 142
Differences between Words with More or Less Familiar Letters………………… 142
Learning Patterns across Blocks in a Multi-modal Task…………………………. 143
**Chapter 6: Conclusion**…………………………………………………………. 144
**References**……………………………………………………………………….. 147

**List of Tables**

**Table 4.1** *One-Way ANOVA Results for Number of Letters, Syllables, and Arabic-only Phonemes*

**Table 5.1** *Participant Demographic and Language Background Information*

**Table 5.2** *Written Stimulus Material from the Letter Training Session and Arabic Word Learning Task*

**Table 5.3** *Means and Standard Deviations of Accuracy and Response Times for the Letter Training*

**Table 5.4** *Mean* Accuracy *and Response Times in the Immediate and Delayed Word Recognition Tasks by Participant Group*

**Table 5.5** *One and Two-Way ANOVA Statistics for Immediate and Delayed Measures*

**Table 5.6** *Two and Three-Way ANOVA Statistics for Immediate and Delayed Word Type Data*

**List of Figures**

**Figure 4.1** *Estimated Marginal Means of Participant Accuracy for Old and New Words in a Match and Mismatch Condition*

**Figure 4.2** *Estimated Marginal Means of Participant Response Times for Old and New Words in a Match and Mismatch Condition*

**Figure 4.3** *Mean Participant Accuracy and Response Times for Old Words Across Blocks 1,2,3 and Session 2*

**Figure 5.1** *Mean Accuracy in the Immediate (Block 3) and Delayed (Block 4) Word Recall Tasks by Participant Group*

**Figure 5.2** *Mean Accuracy in Word Learning for Word Types: all-FAM, un-FAM1, un-FAM2, no-FAM*

**Figure 5.3** *Mean Accuracy in Word Learning by Block and Word Type for Letter Training Groups (LT-High, LT-Low, No-LT)*

**List of Appendices**

APPENDIX 4.1 List of stimulus items and their characteristics used in the 'Multi-modal Arabic word learning in novice L1 English speakers' study

## Chapter 1: Introduction

One of the first challenges learners face in a new language is learning foreign language vocabulary. To establish a lexical representation in memory, an individual requires phonological, semantic, and possibly orthographic information about a word (Laufer, 2005; Nation, 2001; Schmitt, 2008, 2014). Put differently, word learning entails acquisition of the spoken and/or written form of a word and association of that form with meaning (Nation, 2001). Yet decoding and subsequently encoding a word's spoken and written form is not always an easy task for beginner language learners. This is especially the case for native English speakers studying Arabic as a Foreign Language (AFL) (Hansen, 2010; Mohamed, 2016).

In discussing the low outcomes of AFL learners in a task based learning exercise, Mohamed (2016) attributes poor performance to the general difficulties English speakers face learning Arabic vocabulary. There are a number of factors that complicate word learning in Arabic. To begin with, learning the written form of a word requires mastery of the Arabic writing system. For English native speakers, this includes familiarizing themselves with a cursive script which is written from right to left and learning to navigate Arabic's orthography (Abu-Rabia, 2002). Further difficulties arise from the abundance of dialects (Towler, 2017), differences in spoken and written forms (i.e., diglossia) (Ryding, 1991), lack of cognates (England, 2006; Van Mol, 2006), distinctive morphology (Khoury, 2008), and the presence of Arabic only phonemes, or sounds that do not exist in English (Brosh, 2015).

It follows that Arabic is widely considered one of the most difficult world languages for English speakers to learn (Anderson & Suleiman, 2009; Elkhafaifi, 2005a, 2005b; Kuntz, 1996, 2000; Moser, 2013; Omari, 2015; Ryding, 2013). The US Foreign Service Institute rates Arabic a Category V language, estimating it takes English speakers almost four times the amount of study hours to achieve fluency, compared to languages like Dutch and Spanish (FSI, 2020), and studies have found a tendency for high drop-out rates between first and second year AFL learners. For example, Belnap (1987) reported a drop out rate of over 50% in US universities. However, despite this acknowledgement of the heightened challenge level, very little research on AFL vocabulary acquisition exists (Hansen, 2010; Khoury, 2008; Ricks, 2015). This is especially significant given trends in rising enrollment in Arabic programs in the Post 9-11 era and a renewed interest in supporting and advancing the practice of Arabic instruction in the US, the UK, and abroad (Al Batal, 2006; Ramezanzadeh, 2016; Welles, 2002).

As of 2016, a reported 31,554 students were enrolled in AFL programs in Higher Education Institutions in the United States (Lusin & Looney, 2018, p. 16). Enrollment in Arabic has increased significantly over the past 20 years, notably jumping from 10,584 in 2002 to 35,228 in 2009; however, the 7:1 ratio of introductory to advanced level classes remains constant (Lusin & Looney, 2018, p. 17). In the UK, Arabic enrollment at the higher education level is also increasing, but universities report high beginner-level drop-out rates between first and second semesters (Towler, 2018, p. 20). We thus notice a similar trend with relatively few AFL students progressing beyond a beginner level in the language.

In order for beginner language learners to advance, it is necessary for them to build both breadth (i.e., the number of words they know) and depth (i.e., how well they know them) in their receptive and productive vocabularies. In part, this requires effective form-meaning mapping in word learning (Schmitt, 2014). The present thesis is thus part of a wider research agenda that aims to further our understanding of early engagement with Arabic word forms in novice learners. There are many things one can know about a word, from usage-based rules to its part of speech, collocates, and synonyms, but as Webb states "meaning and form is certainly the most important …" (Webb, 2007, p. 65), particularly for beginner learners (Shintani, 2013).

Various approaches to vocabulary instruction can be taken (e.g., presenting spoken and/or written forms, transliterations, translation equivalents, images), nonetheless word learning studies have shown that individuals can acquire foreign language word knowledge in relatively few exposures when multi-modal stimuli are employed (Bisson et al., 2013, 2014a; Cárcamo et al., 2015; Nassaji, 2004; Zarei & Khazaie, 2011). Multi-modal learning entails the presentation of an image and/or audio to accompany the spoken and/or written form of a word. Several studies have used multi-modal vocabulary presentations in an AFL context (Han et al., 2017; Han & Oh, 2018; Showalter & Hayes-Harb, 2015). However, these studies were designed to look at phoneme pattern and phoneme contrast acquisition, as opposed to word learning. Consequently, many of the words began with the same or similar sound and/or letter and this had implications for the diversity of the stimuli set used and the validity of the word learning tasks themselves. Moreover, not all studies involved L1 English speakers.

The primary goal of the first study presented in this thesis is thus to evaluate the use of multi-modal approaches in general AFL vocabulary instruction for native English speakers. A secondary research aim is to discover how the properties of the stimuli (length, number of syllables, number of Arabic only phonemes, frequency, concreteness, image complexity) explain variance in accuracy scores and response times. Despite its less commonly taught language, status, research in AFL is increasingly being used to inform curriculum design and teaching. More insight into difficulty drivers in early word learning can benefit researchers planning AFL studies, teachers, and curriculum developers tasked with designing AFL assessments, lessons, and classroom materials.

One of the most widely acknowledged challenges in early Arabic word learning is the unfamiliar script. The second study presented in this thesis explores the effects of providing the Arabic script to novice learners. No previous learning benefit had been found for absolute beginners who are given Arabic written forms in a learning task (Mathieu, 2016; Showalter & Hayes-Harb, 2015). For this reason, the study introduces a letter training session prior to the multi-modal word learning task. Research inquiries center on the effects of training participants in Arabic grapheme-phoneme correspondences on accuracy and response times in a word-learning task, post-test and delayed post-test measure. Additionally, the study looks at learning outcomes for words containing different amounts of unfamiliar phonemes not included in the training. Optimal repetitions and scope for an Arabic letter training task are also discussed. Findings contribute to our understanding of how novice learners who are given a letter training engage with the Arabic script in early form-meaning mapping and are discussed in terms of their potential to inform and support AFL learning.

Together with the first study, results generally highlight the utility of employing multi-modal word learning approaches, such as those delivered by online platforms and digital apps, in beginner AFL study. These tools have the potential to support classroom learning and could be an effective way of supplementing vocabulary study. In a 2018 survey of language programs offering Arabic in the UK, many schools cited the need for additional resources as a major concern in their delivery of tuition for non-Latin script languages (UCML-AULC, 2018). Results here feed into our understanding of the role multi-modal vocabulary learning platforms can play in advancing Arabic teaching and learning.

**Organization of the Thesis**

Chapter 2 of this thesis presents an overview of the Arabic language and provides background on the unique and nuanced socio-linguistic situation that affects both native speakers and AFL learners alike given Arabic's *diglossic*, or as some researchers prefer, *multiglossic* nature. The discussion of diglossia and its impact on vocabulary learning in AFL is then returned to later in the chapter. Information on Literary Arabic and Spoken Arabic is reviewed, alongside studies that provide context for understanding the challenges researchers face when working with multiple versions of a language, including a lack of shared conventions and abbreviations and a lack of consensus over whether native speakers should be considered bilingual, with their dialect labeled as a first language (L1) and Modern Standard Arabic as a second language (L2). This is followed by a closer examination of the elements commonly cited as factors contributing to the level of difficulty of learning Arabic as a Foreign Language (AFL).

A brief review of Arabic's unique morphology as a Semitic root-based language is undertaken. In turning to the writing system, discussion covers the distinction between *script* and *orthography*. The chapter then presents an extensive review of research with native speakers and AFL learners investigating letter shape differentiation, Arabic's cursive nature, the number of allographs for individual letters, connectedness in the script and vowelized vs. un-vowelized text in relation to visual recognition, decoding, and comprehension in Arabic reading.

Cognates, or words which share form and meaning between Arabic and English, are discussed. Because of script differences, overlap is typically on the phonological level and discussion centers around understanding the learning burden for cognate vs. non-cognate vocabulary in the AFL classroom.

A close examination of differences in Arabic and English phonology is also undertaken, including a review of novel phonemes for English speakers and research on their potential impact on perception of non-native contrasts by L1 English AFL learners. This leads to a discussion of letters with close phonological neighbors, such as the "hard to hear" emphatic consonants, and research regarding how they affect accuracy and word recognition in AFL reading. In surveying available resources for researchers working in AFL, the difficulties of assembling corpora are investigated, including the technical challenges of automatized parsing and morphological analysis in Arabic, navigating and differentiating between heterographic homophones, excluding archaic language forms, including spoken word forms, and developing appropriate selection criteria and weighting of text-types.

The Buckwalter Arabic Morphological Analyzer (2002) and the Buckwalter Arabic Corpus (2010) are covered alongside a review of frequency dictionaries and general concerns about the difficulty of capturing the Arabic language in a frequency list. For example, underrepresentation of high frequency Spoken Arabic labels for everyday items can impact beginner level learning materials and teaching programs that use frequency information to inform content and practice. Attention is briefly given to textbook analysis of *Al-Kitaab* (Brustad et al., 2004), the most widely used textbook in university level AFL programs, as well as a web-delivered assessment tool for AFL (Ricks, 2015) that can be of use to researchers in the field. The chapter ends by reviewing normative image databases for Arabic.

Chapter 3 discusses additional challenges faced by researchers, including difficulty finding participants in sufficient numbers to undertake quantitative research. It then looks at research related to the importance of stimulus modality in word learning studies, namely how a combination of auditory and written input can influence task engagement and affect results. A review of word learning studies involving languages with different writing systems is undertaken and serves to introduce the issue of transliteration and transcription of the Arabic script.

*Transcription* is concerned with capturing the sounds of spoken language whereas *transliteration* is about rendering a language in a set of symbols that correspond to the participants' L1 writing system. The chapter explores how different transliteration systems can capture the nuances of Arabic and the problems posed by conversion of graphemes with no equivalent in Roman script. Additional background is provided on various transliteration systems for Arabic, including Qalam, the transliteration system used in the first study of this thesis. Details are included on appropriate modifications made to Qalam by the researcher to accommodate task constraints in the present research. The chapter concludes with a review and summary of Arabic word learning research involving novice, beginner and intermediate level learners.

Chapter 4 encompasses the first study in this thesis in which British English native speaker participants with no prior knowledge of, or exposure to, Arabic participated in an explicit learning task in which they were exposed to Arabic vocabulary items with transliterated written forms accompanied by images and audio. Models of working memory, the use of images in word learning and literature on multi-modal learning approaches is reviewed, the study's methodology is presented, and results are provided and discussed in the context of learning gains in multi-modal approaches to vocabulary study. Potential difficulty drivers within a stimulus set of Arabic vocabulary items are also covered.

Chapter 5 presents the final study in the present research in which native British English speaker participants undertook a letter training session prior to beginning an explicit multi-modal vocabulary learning task in which vocabulary items were presented in a modified version of the Arabic language script, accompanied by images and audio. A subset of the letter-training group was compared to a control group on immediate and delayed post-tests. Literature concerning letter training is reviewed, the study's methodology is presented, and results are provided and discussed in terms of advantages in learning gains and retention for the group that were trained in Arabic grapheme-phoneme correspondences. Attention is also paid to the results of the letter training session itself, namely why not all participants were included in a high performing subset of the letter-training group, and to difference in performance by the letter-training group on items containing more or fewer graphemes covered in the actual training.

Chapter 6 summarizes the studies' main findings and conclusions, discussing them in the greater context of AFL vocabulary learning.

## Chapter 2: The Arabic Language and Word Learning Studies

### The Arabic Language

Arabic is part of the Central Semitic language family. Like Hebrew and other Semitic alphabets, it is an Abjad language primarily composed of written consonants (Daniels, 2013). Of the 28 Arabic letters, only three, the ا *alif* , و *wāw* and ي *yā'* are vowels but these letters also have dual-function and can sometimes behave like consonants (Ibrahim et al., 2002). For example, the *wāw* can be pronounced as either /oo/ or /w/ and the *yā'* as /y/ or /ee/. The cursive script is written right to left and letters change shape depending on their location in a word. There are also both connecting and non-connecting letters which can result in spaces that occur within a word.

In vowelized texts, a series of diacritic marks appear above and below letters indicating short vowels that occur between consonants. They include the *fatha* /a/, the *kasra* /i/, and the damme /u/. There is also a diacritic for the doubling of a consonant in the case of the *shada* and a lack of a vowel in the case of a *skoon*. The *hamza* ء is often included in the alphabet as a letter but is technically a diacritic that's pronounced as a voiceless glottal stop. Diacritic marks appear in texts written for beginner readers, in texts which have ambiguous words for which form and meaning cannot be inferred from context, and in texts in which precision is required for accurate pronunciation (Dai et al., 2013; Hermena et al., 2015).

Semitic languages are morphologically rich and contain many derived and inflected forms. Arabic words typically have more than one bound or free morpheme (Ryding, 2013) and can minimally be broken down into a root form and a phonological pattern (Perea et al., 2014). For example, the three-letter root ط - ب - خ ***T-b-kh*** means "cooking" or "having to do with cooking." In Arabic the root is also the third person singular past tense form of a verb, thus طبخ 'Tabakha' also means "he cooked." Certain patterns change a root in predictable ways. For example, when the letter *meem* م 'm' precedes the root it changes the meaning to the location where the action takes place, thus مطبخ 'maTbakh' is "kitchen." The person who cooks or a thing that cooks is طابِخ/Taabakha/. The adjective form is مطبوخ /maTbookh/ as in "cooked" food. Note these words look more alike in Arabic orthography than in their English transliterations. This is due to the use in Arabic of diacritic marks instead of written letters to apply patterns to the root forms and may be a factor which complicates the comparison of Arabic word learning studies that use transliterations as opposed to the Arabic script in presentation of written stimuli.

Roots can be made up of two to five letters, but most contain three to four. Arabic has an estimated 10,000 roots and 900 patterns, but not all roots fit all patterns (Al-Sughaiyer & Al-Kharashi, 2004). There is also a set of nouns that do not contain productive roots and there are loan words that come from other languages and cannot be decomposed based on Arabic's morphology. Arabic words can also contain articles, prepositions, pronouns and implied subjects, in the case of verbs. This has led Hansen (2008) to describe them as "information dense" and Ryding (2013) to argue that morphology is even more important than syntax in understanding Arabic.

## Diglossia or Multiglossia?

It is important to explore the diglossic situation that affects the Arabic language as it has implications for teaching, learning and conducting research in the field (Kuntz, 2000; Ramezanzadeh, 2016; Ryding, 1991). Arabic is commonly referred to in terms of its diglossia which Ferguson (1959) describes as a division between the language used by a society for reading and writing and that which is spoken in everyday communication. As a label diglossia suggests a "low" version of the language or colloquial dialect that is used for informal situations and a "high" version of the language reserved for formal and academic communication. However, some researchers (Hary, 1996; Ramezanzadeh, 2017) prefer the term "multiglossia" arguing that it is a more accurate description of the socio-linguistic situation that affects speakers of the Arabic language. This is because Arabic has two literary versions, many distinct dialects, and what the participants in Anderson and Suleiman (2009) referred to as a "third language" that Arabic speakers use to communicate informally outside of their own dialect. Speaking Arabic thus requires the ability to fluently switch between a number of language variants and registers in order to successfully communicate with speakers of one's own dialect and different dialects, in a range of formal and informal settings.

## Literary Arabic

Modern Standard Arabic (MSA), sometimes referred to as Literary Arabic, is the language used by the news media, politicians, and lecturers. MSA is one of the six official languages of the United Nations and is a Lingua Franca for speakers of Arabic dialects living all over the world. While it is primarily a written language, it may be spoken on certain occasions such as during teaching, formal broadcasts, or interactions in which two Arabic speakers have quite distant dialects.

MSA is a descendent of Classical Arabic (CA), the liturgical language of the Qur'an with CA differing mainly in its vocabulary and the greater range of grammatical structures it presents. As opposed to MSA, CA is usually written with the diacritic marks that represent short vowels included, to provide a guide for precise pronunciation (Abu-Rabia, 2001). It is rarely spoken outside of a religious context and may commonly be memorized for recitation, including by speakers of other languages than Arabic who are followers of Islam.

Arabic native speakers grow up speaking their dialect as a mother tongue and don't begin formal study of MSA until the age of six or seven, when they enter school. Beyond exposure to MSA in the classroom, they may also encounter it in documentaries, news broadcasts, or programs created for children who are learning how to read and write (Saiegh-Haddad, 2003).

The research reported in this study makes use of vocabulary taken from Modern Standard Arabic (MSA). This is because MSA is the most common version of the language taught in the Arabic as a Foreign Language (AFL) classroom (Ryding, 1991).

**Spoken Arabic**

Spoken Arabic (SA) dialects can vary from MSA in word order, vocabulary, use of inflected forms for gender, number and tense, and approach to phonology (see Saiegh-Haddad 2003 for a more detailed explanation). Notably, SA follows a subject-verb-object order as opposed to the verb-subject-object of MSA and does not make use of case. SA and MSA also have different but overlapping sound systems. When an MSA word contains a phoneme that is not part of the phonemic inventory of the local dialect, speakers will typically replace it with a native phoneme. For example, in Egyptian Arabic, the letter *jeem* ج is pronounced as a hard /g/ instead of /dʒ/.

Whereas MSA is standard across the Arabic speaking world, spoken dialects fall into different groups: Arabian Peninsula, Mesopotamian, Syro-Lebanese, Levantine, Egyptian, and Maghrebi (Moser, 2013). Both across and within groups they can differ significantly from each other in their lexis, syntax, phonology and use of stress (Nevat et al., 2014). When Arabic native speakers from distant geographies come together, the distance of their spoken dialect to MSA, particularly the transparency of the sound system, can make it easier or harder for them to understand each other. For example, for the North African traveling in the Levant region it will be easier to understand the local dialect than it will be for the locals to understand the traveler, which is where the need for the "third" variety of Arabic that is neither pure MSA nor dialect arises.

### Reporting Dialects in Research

When studies are conducted in an Arabic speaking country, the local dialect is often mentioned (Alduais & Almukhaizeem, 2015). However, when Arabic research is conducted in Western countries, it may be left out. For example, in an eye-tracking study of the behavior of native Arabic speakers in reading a text without diacritic marks (Hermena et al., 2015), no mention was made of the participants' native dialect or country of origin, only that they were living in the UK.

Arabic is the first language of 300 million people in 23 countries (Ramezanzadeh, 2016), and one country may have multiple dialects spoken by people in different regions. Native speakers have been shown to behave in different ways depending on their country of origin. Hansen (2010) used a group of Arabic native speakers from Morocco as a control in her experiment and ended up with surprising results in which the control group was slower in reading an Arabic text than a group of AFL learners. This might have been because a second or third language influenced the linguistic environment in which the native speakers grew up. In Morocco, years of French Colonialism in the North African region resulted in Moroccan children learning French at school at the same time as MSA, and this practice continues even today. Some even go on to complete higher education studies in French vs. the native Arabic.

Hamdi, Ghazali, and Barkat-Defradas (2005) investigated the syllable structure of Spoken Arabic in comparing Moroccan and Tunisian dialects to Lebanese. North African dialects have many long vowels that are either shortened or removed completely from within words. This vowel deletion causes consonant clustering characteristic of Berber, the language of the indigenous people from the region. Lebanese, on the other hand, is much closer in lexis and phonology to MSA. Hamdi et al. (2005) analyzed ten-minute samples of natural speech from Moroccan, Tunisian and Lebanese Arabic native speakers. They looked specifically at intonation, rhythm, and syllable structure, using measures of feet and moras to compare the samples. Their results pointed to a significant difference in the syllabic structure and rhythm of Moroccan and Tunisian dialects as opposed to Lebanese and led them to propose a rhythm continuum on which all Arabic dialects can be placed.

Hamdi et al. (2005) argue that rhythm is fundamental to how speech is recognized and learned. This is one of many reasons why it is important for researchers working in the Arabic as a Foreign Language (AFL) field to both consider and report the dialect spoken by a native-speaker teacher of Arabic, a native speaker who creates a recording for use in a study, or a native speaker who must evaluate or rate a student's performance on oral language tasks, in addition to Arabic native-speaker participants.

In the present study, every effort was made to collect data relevant to the spoken dialects of the individuals providing recordings and the panel of native speakers who assisted the researcher in selecting and describing the Arabic stimuli.

**Talking about Arabic – Conventions and Abbreviations**

Note that there are currently no conventions for how different variants of the Arabic language are discussed in the literature. Modern Standard Arabic and Classical Arabic are both sometimes referred to as *Fusha*, or "the most eloquent" language, which roughly translates to Literary Arabic. This is as opposed to *Ammia* which is an Arabic word used to discuss dialects (Asaad & Eviatar, 2013). The abbreviation CA has been used in reference to Classical Arabic and Colloquial Arabic. Modern Standard Arabic has been discussed using the term Standard Arabic abbreviated as (SA), which consequently can be confused for Spoken Arabic which is also commonly abbreviated as (SA). Ryding (1991) identifies a plethora of terms used to reference the Spoken Arabic used by native speakers communicating outside of their dialect, including Educated Spoken Arabic, Urban Cultivated Arabic, Middle Arabic, Pan-Arabic, Elevated Colloquial and Inter-Arabic, in addition to the Arabic terms *al-luga l-wusTaa* "the middle language" and *lughat al-muthaqqafin* “language of the cultured,” (p. 213). For the research community to work together more effectively, it is recommended that these labels be standardized across all fields, from Arabic Corpus Linguistics to Teaching Arabic as a Foreign Language, to enhance transparency and facilitate inter-disciplinary discussion and collaboration.

**Are All Native Speakers of Arabic Bilingual?**

Native speakers of Arabic are sometimes referred to by researchers as bilinguals in the First Language (L1) literature on literacy skills acquisition (Abu-Rabia, 2002; Saiegh-Haddad, 2003). This is because MSA is structurally, lexically, and phonetically different from Spoken Arabic (SA). In a semantic priming study, Ibrahim and Aharon-Peretz (2005) found larger priming from SA to MSA than MSA to SA and that cross-priming effects for SA words to MSA and Hebrew targets were similar. In bilinguals, forward translation has been shown to be more responsive to priming than backward translation (Basnight-Brown & Altarriba, 2007; Duyck & Warlop, 2009), thus, it was argued that SA is akin to an L1 and MSA a second language or L2. This was also shown by Ibrahim (2006) in a repeated priming study with Hebrew, SA and MSA translation equivalents. However, it should be noted that some backward priming from MSA to SA was found in Ibrahim and Aharon-Peretz (2005). As this backward priming was significant, it points to a more nuanced relationship than that of a traditional L1 and L2 in cross-language priming experiments (Gollan et al., 1997).

Nevat, Khateb, and Prior (2014) undertook an fMRI study that looked at differences in visual stimuli processing and also found results more in line with a diglossic situation. They gave 25 bilingual Arabic-Hebrew participants who had all learned MSA and Hebrew at school, a semantic categorization task with high frequency concrete nouns in which they had to decide if pairs of words in MSA, SA, and the L2 Hebrew were related. They compared activation to a control and found significant effects for language in the four brain regions of interest they had selected. It was predicted that as SA is not typically encountered in written form, it would be treated more like the L2. Results indicated that Hebrew and SA were both slower and less accurate than MSA. However, in a whole-brain analysis SA had stronger activation in many areas in which MSA and Hebrew were more alike. The study thus does not support the idea that SA and MSA are handled like an L1 and L2 in the brain, rather that processing is based on the modality in which the language is more frequently encountered.

**Arabic as a "Difficult" Language**

Arabic is often characterized as a "difficult" language by researchers, teachers and students (Abu-Rabia, 2001; Anderson & Suleiman, 2009; Cicerchia, 2010; Hansen, 2010; Kuntz, 1996, 2000; Omari, 2015; Ryding, 1991) who cite a number of reasons for this assessment including an abundance of inflected and derived terms, the complexity of the script, the vocabulary learning load given diglossia, incomplete phonological information on words in reading, and a lack of cognates with English. It is useful to consider each of these points before discussing Arabic word learning studies and the main issues in Arabic as Foreign Language (AFL) today.

## Morphology

The root-based nature of the lexis means that in Semitic languages, more so than Indo-European ones, there are more words which contain the same letters presented in different patterns (Perea et al., 2014). This feature of the language can be seen as a positive for word learning, as roots and patterns provide insight into the semantic meaning and function of newly encountered terms. For example, Morin (2003) found morphological analysis by L1 English L2 Spanish learners led to productive vocabulary gains, particularly in beginner learners. Some studies with AFL learners (Hansen, 2010; Khoury, 2008) have attempted to test this via a strategy training intervention in root and pattern recognition, followed by measures of word recognition and word formation abilities. These studies will be discussed in greater detail in an upcoming section.

Nonetheless, in deep orthography common roots can also make a language more challenging as they result in words being more easily mistaken for each other, particularly in the case of homographs, words which differ in spelling by only one letter (Hansen, 2008). Additionally, Arabic roots have varying degrees of semantic unity (Ravid, 2005) and not all lexical sets around a root are transparent to learners. Some Arabic words share the same consonants as a root but are not related. Moreover, some Arabic nouns are not derived from verbs (Abuleil & Alsamara, 2004) and other roots only use seven to eight of the 15 possible derivational patterns (Moser, 2013).

## Script and Orthography

In her description of the Arabic writing system, Hansen (2010) provides a helpful definition of terms using *script* to refer to the graphical aspects of writing and *orthography* to the language specific characteristics, including written conventions and spelling.

The script in which the Arabic alphabet is written has several groups of letters that share the same basic shape but are differentiated by the inclusion of dots above and below the letter's x-axis. For example, the ت *tā* , ب *bā* , and ن *nūn* all have close orthographic neighbors (Asaad & Eviatar, 2013). In a study with L1 Hebrew children who were learning Arabic at school, Russak and Fragman (2014) found participants had difficulty differentiating between these letters and often used the wrong form in written production.

Khaldieh (1996) conducted a study with 27 L1 English Arabic as a Foreign Language (AFL) students at different proficiency levels who were studying at an American university. He created a set of stimuli for a matching task in which the correct response contained a phoneme that did not exist in the participants' L1. Each set had both a graphical and a phonological distractor. The AFL students' performance was then compared to that of a group of 9 native speakers. Khaldieh found that beginner and intermediate AFL learners were more likely to choose the graphical vs. phonological distractors, which he interpreted as their experience of difficulty due to the script vs. poor decoding skills. However, he also found a preference for phonological processing in a second sentence level task. This suggests context and the lack of diacritic marks may have affected the reading strategy students used. Of note, there was no mention of the frequency of the stimuli used or the participants' familiarity with the terms and how that might have affected their performance.

The cursive nature of Arabic's script also means that letters change shape depending on if they come at the beginning, in the middle, or at the end of a word, or if they are represented on their own. Ibrahim, Eviatar, and Aharon-Peretz (2002) conducted a study in which they looked at reaction times for native speakers of Arabic who were tasked with analyzing text written in the Arabic script and in their L2 Hebrew. Both languages were presented in un-vowelized conditions and participants were tasked with organizing and connecting a set of randomly scattered stimuli based on the letters they contained. While in Hebrew, five letters change shape depending on their location in a word, in the Arabic script, 22 of the 28 letters have four shapes and the remaining six have two shapes. Results showed that visual recognition reaction times were significantly greater in Arabic than in Hebrew, leading researchers to suggest that the cursive nature of the script and subsequent visual complexity slows its processing, even for native speakers.

Arabic's cursive script contains letters that do and do not connect. When one of the non-connecting letters, for example the ر *rā* /r/ occurs, a word may contain a space that can cause unskilled readers to mistake it for two separate words. Dai et al. (2013) tested the hypothesis that connectedness in script increases difficulty in decoding by examining the effect of letter ligatures with ten native-speaker readers at a private elementary school in Israel. Researchers created two sets of pseudo-words, one in which the words were composed of letters which do not connect and one in which the words were all composed of connecting letters. Words were three and four letters long and participants were exposed to half of the words in each list presented four times in a randomized order. They were given a post-test a week later in which the researchers measured speed and accuracy in reading the words aloud and producing a written version. Results pointed to a significant difference between the reading times for connected vs. non-connected letter words with the non-connected words being read much faster. However, no difference was found in the accuracy of decoding. They also found an effect for familiarity and word-length. Words that had been learned in the previous week were read much faster, as were the shorter words. In measures of spelling accuracy, no familiarity effect was found. However, the difference in accuracy on written production of the connected vs. non-connected letter words was significant. The participants made six times as many errors on connected letter words (Dai et al., 2013).

Asaad and Eviatar (2013) ran a study with four groups of native Arabic speakers, including first graders, third graders, fifth graders and university students. They used a Stroop test combined with a word manipulation in which half of the stimuli violated letter shape rules. For example, they would begin a word with a middle letter position or place a word initial position at the end. They were able to use Stroop because the words for colors are the same in both Spoken Arabic and MSA. Their first-grade participants had the same amount of interference for both sets of words, presumably because they were still learning the skill of decoding in reading. However, the older children and adults experienced more interference in the manipulated letter shape condition as their global reading strategies had been disrupted.

In Arabic orthography short vowels are commonly excluded from texts written for proficient readers. Texts without vowels in Arabic are referred to as having “deep orthography” as opposed to “shallow orthography” because the vowels must be inferred. This is also a feature of the Hebrew language in which a text is referred to as “pointed” or “un-pointed” depending on whether the vowels are written (Abu-Rabia, 2001). Whether a text in Arabic is presented in deep or shallow orthography has been shown to have an impact on accuracy and comprehension in native-speaker reading, with accuracy defined as correctly applying short vowels and pronouncing word endings when MSA is read aloud (Abu-Rabia, 2001, 2002; Dai et al., 2013).

In a study with 65 adult bilingual L1 Arabic L2 Hebrew speakers, Abu-Rabia (2001) presented participants with varying amounts of Arabic and Hebrew text in vowelized and un-vowelized conditions, including single words, paragraphs and a short story. Accuracy in reading aloud and comprehension were measured and a significant difference between conditions within each text-type was found. However, it is interesting to note that in Hermena et al.’s (2015) eye-tracking study of native speaker Arabic readers, there was no significant difference in reading behavior when a text was presented in a fully vowelized sentence vs. un-vowelized condition as compared to a sentence in which only one ambiguous word was presented with diacritics. This suggests that in proficient readers, vowels may be treated as superfluous information and ignored.

Dai, et al.’s (2013) study on letter ligatures also included an experiment that looked at diacritics in reading and how they can increase difficulty in decoding. The researchers exposed 53 native speaker children with a mean age of 8.5 to a series of words with both connected and non-connected letters and varying amounts of diacritic marks representing short vowels. Results showed that there was a significant difference between naming accuracy and speed for words made up of connecting letters with few vs. many diacritics. The more complex the diacritics, the more words were read less accurately and at a slower speed (Dai et al., 2013). This may be relevant in the basic and beginners AFL classroom where vocabulary words are often presented with full diacritics.

Beyond a basic level, AFL learners are often encouraged to read texts without diacritics that Hansen (2008) describes as forcing them to take a whole word approach to word recognition from early on. This means that if the word is unfamiliar, they must guess at the vowels, which can prevent incidental vocabulary learning through reading (Young, 2011). Many SLA studies have looked at the relationship between incidental learning and vocabulary growth in L2 learners (See Huckin & Coady, 1999 for a review). AFL learners are thus at a distinct disadvantage compared to learners of so called "easier" languages, which is particularly relevant given it has been suggested they need to know more vocabulary than students of a European language in order to achieve the same level of comprehension (Van Mol, 2006).

Nonetheless, as part of a series of experiments investigating reading in AFL learners at varying levels of proficiency, Hansen (2010) provided readers with texts in both vowelized and un-vowelized conditions and did not find a comprehension related benefit for shallow orthography. She found significant differences in the vowelized text condition based on proficiency, concluding that learners become better decoders as they move from beginner to intermediate levels when vowels are provided. However, in the un-vowelized text, reading speed was significantly faster across proficiency levels, and no significant differences in performance on a set of comprehension questions was observed between the two conditions. Hansen interprets these results as supporting the view that providing diacritic marks for AFL learners can result in reading becoming a cognitively overloading task (Hansen, 2010). However, she did not have many comprehension questions and the very beginner learners actually read a different text than the other two groups so it would be interesting to try the study with a different group of learners and materials and see if results can be replicated.

**Diglossia**

Learning Arabic is complicated by the number of variants that exist – see previous section on Diglossia or multiglossia for a review. Moser (2013) explains "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." (p. 20).

For those students who undertake simultaneous study of a dialect and MSA, there is an additional word learning burden where lexical inventories differ. While older publications have suggested there might be double the amount of vocabulary to learn than in other foreign languages where spoken and written forms are closer (Al-Batal, 1992), a lack of empirical research to support these claims means the extent of the difference is still unclear. Young (2011) underscores that difficulty arises from the fact that students must master synonyms, also known as *lexical doublets*. Differences in SA and MSA exist regardless of word-category, for example the time-marker "yesterday," the noun "breakfast," and the verb "to speak" must all be learned in both variants (Moser, 2013, p. 17). The pairs also need to be produced in the correct context, meaning students must master the complex socio-cultural rules tied to which version of the language to use and when. This is particularly difficult at a beginner level when basic skills such as decoding and pronunciation are still being honed. When AFL learners are only familiar with one form of the language, they may experience heightened anxiety in situations where the other is used, which can detract from learning and affect performance in listening and speaking (Elkhafaifi, 2005b).

Even for those individuals who have been exposed to a spoken variant of the language, such as heritage learners whose family members spoke an Arabic dialect at home (Ramezanzadeh, 2017), diglossia adds to the difficulty in learning MSA. For example, Saiegh-Haddad (2003) showed that in native-speakers, differences in the phonemic inventory of SA and MSA had an effect on comprehension, short-term memory, reading and sensitivity to MSA sounds. Of note, the ق *qaf* /q/, ة ta marbuta /ta/, ث *thā* /th/, and ظ *zā'* /Z/ are not part of the phonemic inventories of most spoken dialects (Asaad & Eviatar, 2013).

**Cognates and Phonology**

Arabic and English are quite distant languages in terms of morphology and script, but they also have very little overlap in their lexis and some major differences in their phonemic inventories.

Cognates are words that are similar in form and meaning (Voga & Grainger, 2007). In Arabic and English form overlap typically occurs at the phonological level, given differences in script and reading direction (Gollan et al., 1997). Cognates may exist due to language contact or because of shared roots, which is more frequently the case between languages from the same family. Van Mol (2006) estimates that only one out of every 1,000 words in Arabic is a cognate in English. Note that the same cognate ratio may not be true for other languages, particularly those in which Arabic has provided a number of loan words including Turkish, Farsi, Urdu, Dari, Pashto and even Spanish.

In comparing Semitic languages like Arabic and Hebrew to Indo-European languages, cognates are often loan words. They may be loan words from English into Arabic but there are also some Arabic loan words in English, such as *"alcohol"* and "*yacht*." Khrisat and Mohamad (2014) posit that cognates are more commonly encountered in SA than in MSA.

Differences in the degree to which cognates resemble each other in terms of phonology, orthography, and meaning, has led some researchers to suggest that they exist on a continuum of similarity (Allen, 2013). In Arabic and English the lack of a shared script, differences in the phonemic inventories of the two languages, variations in use of stress by speakers of different Arabic dialects, and the practice of providing written MSA without short vowels may all affect the degree to which cognates in the L1 and L2 are recognizable to AFL learners (Van Mol, 2006).

Khoury (2008) calls for more research to be done in this area to fully understand the learning burden for cognate vs. non-cognate vocabulary words in the AFL classroom. From a research perspective, it would also be interesting to build on previous priming studies (Gollan et al., 1997; Ibrahim, 2006; Voga & Grainger, 2007) to further investigate if the opaqueness of the phonological and/or orthographic form of Arabic cognates prevents some AFL learners from recognizing them and experiencing a word-learning benefit in AFL word learning.

In terms of phonology, Arabic contains more consonant phonemes than English -- 28 as compared to 24 (see Al Mahmoud 2013 for a review), including some that are non-native for English speakers. These include uvular and pharyngeal consonants, the glottal stop, and emphatics (Hayes-Harb & Durham 2016). Of note, the pharyngeal sound corresponding to the letter ع *'aiyn* /'/ is often perceived by learners and teachers as particularly difficult for English speakers (Cicerchia, 2010) and some researchers have used the presence of non-native phonemes to assign word difficulty to stimuli in AFL word learning tasks (Golonka et al., 2015).

However, the status of phonemes as native or non-native does not necessarily equate to higher levels of difficulty in producing them or differentiating between them (Best et al., 1988). Lababidi and Park (2014) asked English native speakers with no previous exposure to Arabic to label Arabic consonants using English phonemes along with an evaluation of the quality of the fit for the sound pairing, to produce an index of Arabic to English pairing. Al Mahmoud (2013) investigated the perception of non-native contrasts in L1 English L2 Arabic North American AFL learners. He gave 22 university students, none of whom had spent more than six months living abroad in an Arabic speaking country, a series of discrimination tasks. A native speaker recorded sets of words and participants wrote down whether the first or third word was the same as the second word. He recorded results for all the non-native phonemes (See Al Mahmoud 2013 for a review) and found that their written forms caused interference as participants confused certain letter forms.

The Arabic alphabet contains letters with close phonological neighbors, notably the emphatics, for example ط /T/ and ت /t/, ض /D/ and د /d/, س /s/ and ص /S/, ظ /Z/and ث /th/ and the ك /k/ and ق /q/. There is also a set of letters that have both visual and phonological neighbors. In a series of experiments with L1 Hebrew children learning

Arabic, Russak and Fragman (2014), and Fragman and Russak (2009) showed that words containing orthographic neighbor letters and letters with both phonological and visual neighbors were more likely to result in spelling errors in productive tasks.

Hayes-Harb and Durham (2016) investigated hard to hear consonants in AFL learners and found that a speaker's perception of an emphatic was influenced by the vowel that followed it. There are fewer vowels in Arabic than in English, but there is also a distinction made between long and short vowels. Like Russak and Fragman (2014) this long vs. short contrast has been cited as the source of spelling errors on vocabulary tests in L1 English AFL learners in which a long vowel was mistaken for a diacritic and left out, or a long vowel was erroneously added to a word (Cicerchia, 2010; Khaldieh, 1996).

In a series of experiments that looked at how the Arabic script affects AFL learners' word recognition abilities in reading, Hansen (2010) ran an experiment in which she measured accuracy and speed in a decoding task in which three types of texts were used. One was composed of 39 pseudo-words constructed following the rules of Arabic morphology and excluding any Arabic-specific phonemes, the second text was the same except pseudo-words included these Arabic specific phonemes, and the third was a Romanized transcription that excluded phonemes that have been identified in the research as problematic for AFL learners (Khaldieh, 1996). Participants were 71 AFL students at three different levels of proficiency and from various L1 backgrounds including Danish, English and German, and 24 Arabic native speakers. They had to read the stories out loud as quickly and as accurately as possible with the researcher tracking and coding various types of errors. No effect for text-type was found in the results for the AFL learners. However, unexpectedly, the native-speaker readers were significantly slower in reading the pseudo-words that contained the native only phonemes. This is significant as the native speaker group were all participants from Morocco whose native dialect is seen as being one of the furthest from MSA, in terms of structure, lexis and phonemic inventory. They are also more likely to have had ample exposure to Romanized script due to the influence of French on the local education system.

While there has been some recent work in this area (Al Mahmoud, 2013; Fragman & Russak, 2009; Hansen, 2010; Hayes-Harb & Cheng, 2016; Lababidi & Park, 2014; Russak & Fragman, 2014), more research is needed to understand how Arabic phonology affects difficulty in AFL word learning, particularly in L1 English learners.

**Resources for Researchers**

**Arabic Corpora**

As a less commonly taught language, it is expected that there will be fewer resources available for teachers and learners of Arabic as compared to European languages, but this dearth of materials also extends to individuals who wish to work with Arabic as a research language (Moser, 2013; Ricks, 2015).

Al-Sulaiti and Atwell (2006) pointed out that while Arabic has a rival number of native speakers to English, no general or reference corpora existed on the same scale as major English projects like the 100,000,000 words of the British National Corpus (*The British National Corpus*, 2007) in the early 2000s. Most of the available corpora were either not free, not tagged for part of speech, or were specialized corpora that focused on a particular text-type or language variant (see Al-Sulaiti & Atwell, 2006 for a review). For example, there was a corpus made up of passages from the Holy Qu'ran and another of the spoken Arabic dialect from Tunisia (Al-Sulaiti, 2010).

Corpora, defined by Cheng (2012) as collections of texts assembled based on sets of design criteria, are essential tools for language researchers. They can provide insight into lexis, collocations, syntax, and language use and are the basis for estimating frequency information for vocabulary. While today there are more large-scale and public-facing projects like Camel Treebank, an open-source dependency treebank housing 13+ corpora for Modern Standard and Classical Arabic (Habash et al., June 2022) and the International Corpus of Arabic (*International Corpus of Arabic*, 2013) at 100,000,000 words, Ricks (2015) explains that "Arabic corpus linguistics is still in its early stages, and faces some striking challenges," (p. 4). The development of Arabic corpora has been held back by a number of factors. These include the challenge of assembling a representative sample of text in light of Arabic's diglossic nature and the creation of tokenizers, morphological analysis engines, and part of speech taggers that can effectively mine corpora to render them useful tools for language researchers, learners, and instructors (Abuleil & Alsamara, 2004; Al-Sulaiti & Atwell, 2006; Ricks, 2015).

In turning first to the problem of which texts to include in a corpus, Al-Sulaiti and Atwell (2006) developed a survey tool to question AFL teachers and other stakeholders who might serve to benefit from the corpus, about their needs. They used the responses to develop selection criteria for texts and assemble a corpus of 100 texts of 100-4,000 words including magazines, website content, newspapers, radio, and email text available online, entitled the Corpus of Contemporary Arabic, which today contributes 1,000,000 words to the Leeds Internet Corpora. The weighting and text-types included in the CCA is indicative of the situation in other “general” corpora, which include mainly MSA from media-based sources. Additionally, there is the problem of archaic word-forms entering frequency data due to the presence of Classical Arabic in the corpora’s texts (Attia et al., 2011). Spoken language tends to be less represented in corpora because of the difficulty in collecting and transcribing it (Cheng, 2012), and this problem is even more exaggerated in Arabic (Ricks, 2015) due to the high number of dialects and the difficulties of undertaking transcription in Arabic, which will be discussed in greater detail in an upcoming section.

In what regards the technical challenges associated with automatized parsing and morphological analysis, it is useful to first review the processes undertaken in English corpora. English words are commonly subjected to stemming (Al-Sughaiyer & Al-Kharashi, 2004), which is when their suffixes or affixes are paired back to reveal a stem form. This works for English because it is a concatenative language in which word parts build on each other in a linear fashion to create meaning.

However, Arabic is a morphologically complex language in which words are made up of a root form paired with a derivational pattern. Words may also be systematically inflected and have articles and pronouns attached (Ryding, 2013). This adding on of different parts of speech to Arabic words, in addition to the presence of connecting and non-connecting letters in the cursive script, makes it harder to use a standard space measure for word demarcation (Al-Sughaiyer & Al-Kharashi, 2004). Reading text from images is also more difficult due to a lack of consistency in whitespace use in handwritten words, archaic font types, and widespread misuse of certain letters by native speakers (AbdelRaouf et al., 2010).

Additionally, heterographic homophones abound in Arabic (Abu-Rabia, 2001). These are words that may be spelled in the same way when no diacritic marks representing short vowels are present, which is the case for the majority of Arabic texts. An automatized engine needs to be highly sensitive to context in order to differentiate between them with the same degree of accuracy as a native-speaker reader. Often human involvement is required, which makes many Arabic corpora projects cost prohibitive (Ricks, 2015).

A number of approaches have been taken to resolve the challenge of identifying word boundaries and extracting root forms in Arabic corpora (see Al-Sughaiyer & Al-Kharashi, 2004 for a review). Most are based on a multi-step process in which suffixes and affixes are stripped and words are compared to tables of roots and patterns in search of a match. As such, foreign loan words and proper nouns that do not conform to Arabic morphology can be problematic for these systems (Atwell et al., 2004).

The Buckwalter Arabic Morphological Analyzer (2002) now called the Standard Arabic Morphological Analyzer (Attia et al., 2011) is used for look-up and part of speech tagging by the majority of corpora today and includes tokenization, word segmentation, dictionary look-up, compatibility checks, and a second look-up for orthographic variants. It was used to mine the Buckwalter Arabic Corpus, a privately held corpus of 2.5-3 billion words, and led to the publication in 2010 of Arabic's first modern frequency dictionary, *A Frequency Dictionary of Arabic: Core Vocabulary for Learners* (Buckwalter & Parkinson, 2011).

**Frequency Dictionaries**

While Buckwalter & Parkinson (2010) is an improvement on early frequency lists such as Landau's *A Word Count of Modern Arabic Prose* (1959) based on the 136,000-word corpus *The Basic Word List of the Arabic Daily Newspaper* (Brill, 1940) there is still room for improvement in assessing what Ricks (2015) refers to as Arabic's "lexical shape," (Cobb & Horst, 2004).

In a thesis on teaching and learning Arabic vocabulary, Young (2011) focused on the fact that the Buckwalter corpora gave more weighting to Media Arabic and online MSA sources as texts, which resulted in a different picture of frequency in Arabic as compared to other languages. Prepositions and politically relevant words may be listed as high frequency whereas labels for everyday items, which are more commonly referred to using spoken dialects, are underrepresented. This is of course an issue with diglossia and vocabulary instruction in general, but is a key point for teachers who use frequency information to inform their practice.

Words that may be listed as high frequency in English, such as جبن "*cheese*" from the 2000-3000 band, appear in the 5,000+ frequency range in MSA according to Buckwalter and Parkinson (2011). AbdelRaouf et al. (2010) took this into account in the development of their Multi-modal Arabic Corpus (MMCAC) purposefully including language from everyday life that appeared in both image and text-based sources.

Familiar's recent publication *A Frequency Dictionary of Contemporary Arabic Fiction: Core Vocabulary for Learners and Material Developers* (2021) also addressed this issue by creating thematically organized frequency-ranked lists for topics likely to be of use to learners, including food, emotions and places. Familiar drew frequency data from 144 literary works of Arabic fiction.

Returning to Buckwalter and Parkinson (2011), it's also important to note they used lemmas instead of morphological variants to determine word frequency, combining past and plural forms of verbs into one lemma and singular and plural nouns into another. This is problematic when you consider Arabic plurals are highly irregular and must be derived. Moreover, while lemma-based noun frequency data is useful for beginner learners in English, it may be less so in Arabic given the opacity of some derived forms and the difficulty beginner learners can experience in extracting roots and noticing patterns (Olshtain, 1987).

The challenge in accurately capturing the Arabic language in a frequency list is also acknowledged by Oweini and Hazoury (2010) who undertook a project to compile a sight word list for native speaker children developing reading fluency. Instead of using available frequency information, they elected to hand-pick items to introduce the basics of Classical Arabic phonology and speed the transition to reading texts without diacritic marks. In their study of word processing by native-speaker bilinguals, Nevat et al. (2014) similarly elected not to use an established frequency list and instead gathered frequency information for both SA and MSA stimuli using a group of native-speaker raters. Presumably the frequency data they obtained more accurately reflected Arabic language use in the specific geo-political environment in which their participants lived.

## Text-book Analysis and Assessment Tools

With the publication of a frequency dictionary came the creation of new frequency informed resources that had not previously been available for teachers and learners of Arabic as a Foreign Language. Moser (2013) undertook an analysis of the most widely used textbook in the UK and US in university level Modern Standard Arabic instruction, evaluating the taught vocabulary in *Al-Kitaab* (Brustad et al., 2004) and including frequency as a measure. This research is of great value to anyone designing word-learning experiments with AFL students who are likely to have used *Al-Kitaab* at some point during their studies. While much of the vocabulary presented could be found in Buckwalter and Parkinson (2011), Moser (2013) reported a gulf between the needs of learners in terms of language use and the lexis presented.

Similarly, Ricks (2015) used frequency data to inform the creation of a web-delivered assessment tool for AFL learners. He piloted and validated his testing tool with 161 AFL learners at varying levels of proficiency. He also measured the effect of frequency on item performance – finding a correlation between frequency and item difficulty ratings. He discusses the fact that assessing proficiency levels using a validated research instrument is common practice for researchers working with English language learners. However, no such tool was previously available for teachers and researchers working with Arabic.

## Normative Image Databases for Arabic

English language researchers have assembled large data sets like Snodgrass and Vanderwart's (1980) bank of 260 line drawings, which includes information on name agreement, image agreement, age of acquisition and visual complexity for a set of concrete nouns. These databases are often translated into major European languages (Desrochers & Thompson, 2009). A recent project run by the Basque Center on Cognition, Brain and Language (Duñabeitia, 2017) has created a set of 750 color images and requested normed versions be created for additional languages. However, when it comes to similar resources for Arabic, up until a few years ago there was not much to be found (Boukadi et al., 2016).

When images are used in a study, they might be considered as an additional variable because participants can have difficulty understanding their meaning or may interpret them in unique ways. Thus, making use of a standard set of images that comes with normative data and applies to the target language group you are working with, improves the quality of the research and results. It also makes it easier to compare data from different experiments.

In recent years, several researchers have attempted to create both new image sets and Arabic versions of existing sets (Alyahya & Druks, 2016; Boukadi et al., 2016; Khwaileh et al., 2014). However, many of the images used in existing sets, for example the American football helmet and the sled in Snodgrass and Vanderwart (1980), were either not culturally relevant in an Arabic language context or not appropriate. Some of the animals in Masterson, Druks, and Galienne (2008) are non-native species in the Middle East that are not well known, consider moose and armadillo. The same phenomenon can be observed in a set of action pictures (Masterson & Druks, 1998) that include images of skiing and skating. These are not common sports in the Arab world, and their translation equivalents are used virtually interchangeably by native speakers of Arabic.

Boukadi, Zouaidi, and Wilson (2016) cite issues with the cultural appropriateness of some of the images in their object battery adaptation for Tunisian Arabic. They started with a total of 400 words and ended up with 348 in the final study. Of note, because they were creating norms based on a spoken dialect, they collected subjective measures of frequency and their word naming responses were recorded vs. written by participants (Bokadi et al, 2016). In their adaptation of the Object and Naming Battery into Saudi Arabic, Alhahya and Druks (2016) created an Arabic version of the battery removing any images which failed to achieve high naming agreement, for example *hammock*, any images which were named with a semantically related word, for example *berries* in the place of *cherry*, alongside any images which did not receive a response, were different in dialects, or were culturally irrelevant.

Khawaileh, Body, and Herbert (2014) set about to create a normative database for Levantine Arabic nouns selecting images that corresponded to masculine and feminine words, broken and sound plurals, and rational, as in referring to human beings and irrational, as in referring to non-human beings, nouns in an attempt to provide increased representation of Arabic-specific language forms in an image database. Arabic plurals are irregular, and some forms split the root consonant cluster using a long vowel, which results in a broken form, as compared to words which have plural endings added on. Their final set included 186 images of nouns for which they collected visual complexity, imageability, age of acquisition and naming latency and agreement data from 22 native speakers. However, the dataset is of limited utility for a number of reasons. First, it has a very small number of items. Second, while participants were native Arabic speakers and their country of origin was reported, they were all living in England at the time of the study. Third, the images themselves were photographs taken from copyright free websites, thus they were not standardized.

## Chapter 3: Arabic Word Learning Research

### Additional Challenges

Part of the problem in conducting research in Arabic as a Foreign Language (AFL) is the lack of adequate tools available to researchers, as discussed in the previous chapter. However, there is also the difficulty of finding participants in sufficient numbers to undertake quantitative work. According to some researchers, this may be why the AFL learner population is such an infrequently researched group (Kuntz, 2000; Maghrabi, 1997).

The following sections will review additional difficulties related to presentation of stimuli and selecting a transliteration system for stimuli that are not provided in the Arabic script, before beginning a review of beginner and intermediate level AFL research.

### Modality of Input

Stimuli can be delivered in various modalities, but research has shown that both the choice of modality and the combinations presented can affect the way participants engage with a task. Detey and Nespoulous (2008) observed syllable segmentation in Japanese L1 learners who were given a set of French pseudo-words under three conditions: auditory input only, visual input only, and audio and visual stimuli presented together. French is a relatively opaque language in that words may contain a number of letters that are not pronounced. Results showed that the Japanese participants were sensitive to the orthographic L2 forms: the auditory only group was less prone to introducing new sounds than individuals in the visual and audiovisual conditions.

In Escudero, Hayes-Harb, and Mitterer (2008) Dutch L1 participants who received both written and auditory input were better able to identify word-initial syllables in English than those who had only heard an audio recording. Escudero (2015) showed orthographic input was sometimes used as an additional aid for L1 Spanish and English participants distinguishing between minimal pairs.

Hayes-Harb and Cheng (2016) ran a study with L1 English speakers that looked at acquisition of Mandarin words presented in either a Zhuyin or Pinyin writing system. Zhuyin is made up of unfamiliar symbols whereas Pinyin is written in a Latin script. Results suggested that the unfamiliar graphemes in Zhuyin took twice as many exposures for participants to learn as compared to the Pinyin Romanized script group. Nonetheless, the Pinyin group was more influenced by their L1's phonology in selecting foils in a word-learning task.

**Transliteration and Transcription**

In preparing Arabic text for Western readers, transliteration and transcription methods are often undertaken, thus, it is useful to include a brief discussion of related terms (Ryding, 2013). *Transcription* is concerned with capturing the sounds of spoken language whereas *transliteration* is the process of systematically rendering a language in a set of symbols that correspond to a different writing system. Arabic transcriptions can be done in the target language script, which is the case when spoken dialects are written using Arabic orthography, in the international phonetic alphabet, or in a Romanized script, which may cause them to be confused for transliterations.

As discussed in the section on Arabic corpora studies, the root system, coupled with the absence of written short vowels, has resulted in the language having a large number of heterophonic homographs. These are words that vary in their pronunciation and meaning but have the same form when a text is presented in deep orthography. In some cases diacritics are provided for ambiguous words to facilitate comprehension in native speaker readers (Hermena et al., 2015) but more commonly a heterophonic homograph's identity must be inferred from context. For this reason, expert knowledge is required for accurate transliteration and transcription to identify words and supply the correct diacritic marks.

Consider the many ways in which Arabic proper nouns are written in English (Fattah & Ren, 2008). You may encounter *Muhammad*, *Mohammed*, *Mohammad*, *Mohamed* or *Mohumed*. In Arabic, the name is commonly written with four consonants that map to an English 'm-H-m-d.' The various forms of the name reflect attempts to transcribe the short vowels between the letters and account for the diacritic marks that inform the name's pronunciation.

Additional concerns in transliteration center around the conversion of graphemes which have no equivalent in Roman script. This is the case for Arabic emphatics (Hayes-Harb & Durham, 2016). Problematic letters in Arabic transliteration also include the *'aiyn* and the *hamza*. Rendering long vowels like the *alif*, *yā'* and *wāw*, which can vary in their pronunciation adds a further challenge. An optimal system must find a way to accommodate the novel forms using existing symbols or combinations of symbols with as close to a 1:1 mapping as possible.

Lawson (2010) undertook an evaluative study of the different systems used by national libraries in the United States in an attempt to enhance usability for individuals looking for materials written by Arabic speaking authors. His initial research resulted in what he referred to as a “substantial list of transliteration options with no one system particularly dominant in terms of overall prevalence,” (Lawson, 2010 p. 165). As such he selected six systems and created a scoring rubric in order to compare them based on metrics including adherence to pronunciation of the Arabic letter and use of non-native diacritics in the transliterated version.

The intended use of a transliterated form will, to some extent, influence the system selected. For example, if the reader is meant to learn a word’s spoken form a more accessible transliteration that provides phonetic clues may be appropriate. If the end user is a computer that needs to accurately manipulate words in a database, it would be better to have unique letter to letter mapping.

Qalam is morphologically based and can be used by computers to transliterate text back and forth between Latin and Arabic-script languages. It handles the emphatics by way of capital letters, which is less faithful to the original phonology but may be easier for English L1 readers than dealing with non-native diacritics. SATTS is a Standard Arabic Transliteration System used by the military and containing a 1:1 mapping with English letters at the expense of accurate phonetic rendering. Buckwalter was developed by the Xerox corporation and used in early automatized systems. As it is concerned with maintaining accuracy in automated transliteration, it is not always a close rendering of the original phonology.

In reviewing the literature, some studies which utilize Arabic transcriptions and transliterations do not mention the actual transliteration or transcription system used (Fitzpatrick et al., 2008; Showalter & Hayes-Harb, 2015; Tseng et al., 2018). This makes it harder to compare their results. Arabic is a particularly challenging language, both when it comes to selecting a transliteration or transcription system and undertaking the actual conversion (Fattah & Ren, 2008; Ryding, 2013). Casual approaches to Arabic transcription can result in a greater variance of spellings and consequently affect a study's written input.

Stevens (1996) recommends presenting a phonetic transcription of vocabulary words to beginner AFL learners and Hansen (2010) found that AFL learners who were familiar with the Arabic alphabet read a Romanized transcription faster. In Tseng, Doppelt and Tokowicz (2018) 32 L1 English participants with no prior knowledge of Arabic learned 51 words and 34 phrase translation pairs over seven sessions. Training was either with or without transliteration. The researchers hypothesized that transliterations would clarify sounds that may have been unfamiliar and help participants better remember spoken forms, generally supporting Arabic vocabulary learning. Participants heard an English word and its Arabic translation twice, once fast and once slow. In the transliteration condition the word was also provided at the bottom with its English definition at the top. Results showed the transliteration presentation led to significantly better free recall, greater accuracy, and faster reaction times on a translation task. Researchers also looked at working memory and reported that participants with lower working memory benefited more from the transliterations.

Due to the relative inaccessibility of the language by English speakers who are not familiar with the Arabic script, the first study presented in this thesis makes use of transliterated versions of Arabic written forms using the Qalam transliteration system (Heddaya, 1985). This is due to Qalam's high usability scorings and popularity of use in different mediums. However, several systematic phonetic adjustments were made to clarify the pronunciation of dual-function letters that can be pronounced both as long vowels and consonants, and to reduce noise in the transliterated versions, particularly in the form of MSA case endings which are not pronounced. This was done in an attempt to make it easier for an English speaker to read these words and renders the transliteration more straightforward.

**Word Learning Studies**

The section that follows presents a review of additional work that has been done with Arabic word learning, concentrating primarily on studies that involve low to low-intermediate level learners.

In a single-participant case-study, Fitzpatrick, Al-Qarni and Meara (2008) looked at the list-learning method of vocabulary acquisition using a set of 300 high-frequency Arabic words. An L1 English speaker with no prior exposure to Arabic who teaches linguistics at the university where the study took place learned 15 words a day over the course of 20 days using a set of flashcards which contained the English transcription of the word, and the English translation along with a place for taking notes regarding the days on which a given word was revised. Researchers administered a series of receptive and productive tests at the end of the last learning session and then at two, six and 10 weeks following the conclusion of the learning period. The immediate post-learning period test measure showed the participant was able to learn 286 of the words. Those acquired in early learning sessions were retained better than the more recently learned words as they had more opportunities to come up in revision, and receptive knowledge experienced less attrition than productive knowledge. This is in line with previous studies on word-learning and the importance of repetition (Ellis & Beaton, 1993; Laufer, 1997).

Notwithstanding the limitations of a single participant case study, the study does not mention how the words were selected or transcribed. It also does not include a description of the words in terms of length, part of speech, concreteness or imageability. The words themselves are also not provided as an appendix. Additionally, Fitzpatrick et al. (2008) reported an effect for word length, concluding that shorter words were easier to learn than longer words, which is supported in the SLA literature (Laufer, 1997; Nation, 1990) but is not supported by any direct evidence reported in the paper.

In analyzing performance on post-tests against a learning diary kept by the participant, Fitzpatrick et al. (2008) reported lower rates of attrition for words for which the participant had created a mnemonic device. The beneficial effect of mnemonic devices which use phonologically linked keywords from the L1 and mental imagery to create an L1-L2 semantic link was also shown in a qualitative intervention study conducted with beginner AFL university learners by Cicerchia (2010). The researcher trained a class of L1 English L2 Arabic learners in in the Keyword Method and then measured performance in a series of productive and receptive vocabulary learning tasks, reporting accuracy rates and analyzing interviews with learners discussing their approach to vocabulary learning in Arabic.

For her doctoral dissertation on vocabulary acquisition in AFL, Khoury (2008) worked with what she calls "high-beginner" AFL learners who had received eight months of one hour/day, four days a week study of Arabic at the time of the intervention to conduct a root and pattern word-learning study.

It has been suggested that the memorization burden in word learning is reduced if a learner both notices a word's root and is able to make generalizations about forming new words using the same root embedded in different patterns (Al-Batal, 1992; Stowasser, 1981). For example, Olshtain (1987) found that morphological awareness positively impacted production task innovations in Hebrew L2 learners. Ravid (2005) also studied root perception in Hebrew L1 and Arabic L1 children and adults and cited the semantic unity of a root and its phonological opacity as important factors impacting performance.

Khoury's study (2008) used 56 students from three AFL classes who were randomly assigned to a control or treatment group. In the treatment group participants were provided with a PowerPoint presentation in English demonstrating word formation rules in Arabic, along with practice worksheets that presented six verbs and 12 nouns that were not found in the students' textbook and had not been identified as familiar on a pre-test. Pattern formation was practiced and a pen and paper test was given after the learning session, along with a delayed post-test. Stimuli included two 3-letter root words for six possible patterns. The control group also received exposure to vocabulary acquisition training in the form of a flashcard strategy.

Results from the post-tests showed the treatment group excelled in word innovation as compared to the control. This is not surprising given it was the skill they were trained in. Khoury (2008) also reported that the words that were subjected to morphological analysis were less likely to be forgotten by the treatment group in the delayed post-tests. However, no significant difference was found in their ability to infer the meaning of new words based on familiar roots or patterns.

Nonetheless, caution is needed in interpreting these results in the context of the beginner AFL classroom. Khoury (2008) included three classes with “comparable proficiency rates” yet no individual measure of proficiency was taken, and student abilities can range significantly within a class. Moreover, in Khoury’s (2008) pilot study with students who had only one to three months of exposure to Arabic, the control group outperformed the treatment group. This may have been due to the lower proficiency level of the learners, particularly their lesser ability in decoding the Arabic script, or it could have been related to a difference in materials. In her main study Khoury (2008) taught fewer words, roots, and patterns.

Khoury (2008) recommends that for the root and pattern strategy to work, learners need to have multiple exposures to a word, yet how many and in what input mode (written, oral or written with oral) is an area for future research.

It is also worth mentioning that Hansen (2010) looked at root and patterns knowledge in a previously mentioned study involving AFL learners at varying levels of proficiency. She created nine pseudo-words that could only accommodate one or two different vowel patterns and had participants read them aloud. Beginner learners made significantly more errors than advanced learners who still did not achieve the accuracy of a native speaker control group.

In what is likely the most comprehensive vocabulary acquisition study conducted with AFL learners, Golonka et al. (2015) looked at the role of context and cognitive effort in vocabulary learning in intermediate AFL learners with the goal of identifying tasks that can enhance lexical gains in three conditions: decontextualized, semi-contextualized, and fully-contextualized vocabulary learning. Participants were 91 L1 English AFL learners from an intensive US Government foreign language course who took part in the study as part of their normal school day.

Buckwalter and Parkinson's (2011) list of the 5,000 most frequent words in MSA was used as an initial source for the stimuli with words in the top 1,000 most frequent band removed, as well as any item the students might have been familiar with from their textbook, words which were not nouns and those terms for which an obvious root could not be identified. They also controlled for word length, pronunciation difficulty, which was determined by the number of non-English phonemes in a word, number of syllables, concreteness, and imageability of the word, in order to create six separate word lists that were randomly assigned to participants.

Researchers developed 12 tasks for the various conditions. Tasks ranged from those requiring low to high effort and included reading and answering true/false questions, filling in the blanks, reading and writing a gist in English, identifying word roots, deriving meanings using new roots, unscrambling sentences, studying word pairs, semantic categorization, and recall practice. They assigned different cognitive effort values to the various tasks and introduced words gradually over the course of multiple sessions. Participants had six, 15-45 minute sessions once per week and were aware that they would be tested at three points: immediately following the session, one week later and several weeks after the session. Their attention was drawn to the target words via glossing in English in the margin of the fully-contextualized condition, presenting a sentence along with its English translation in the semi-contextualized condition, and presenting Arabic and English translation equivalents in the de-contextualized tasks.

Two raters with a high inter-rater reliability score evaluated performance on the tasks. Researchers gave the Vocabulary Knowledge Scale (VKS) (Wesche & Paribakht, 1996) at the conclusion of each session which also included a self-report measure. Results were analyzed with mixed-effect modeling and showed that word recall was significantly better for words presented in high vs. low-effort tasks. Accuracy was greater when words were presented on their own and decreased significantly for words learned in tasks with semi and full-contextualization. Only the full-contextualized condition showed a significantly different recall result between the immediate and delayed post-tests.

While the large range of tasks involved in the study suggests the importance of engagement and using high-effort tasks in the AFL classroom, the researchers did not look at the properties of the Arabic words in relation to recall success. It would be interesting to see if there was an effect for word length, frequency, imageability and concreteness both in immediate and delayed gains and self-reports in the VKS post-tests.

**Summary of Research Reviewed**

In looking at the studies reviewed here together, a common issue is a lack of pre-test measures in attesting to the proficiency of the AFL learners who took part. Participants are typically categorized based on hours of exposure in class versus any standardized assessment measure, and thus it makes it difficult to compare or generalize results to a greater AFL population. This is even more relevant when you consider AFL classes often contain heritage learners. Ramezanzadeh (2016) points out that within heritage learners there exists considerable differences between participants' familiarity with and everyday exposure to the language, which is reflected in performance on standardized UK language exams.

This is one reason why the present research aimed to remove any ambiguity by working only with complete beginners who had no prior exposure to Arabic and were not heritage learners.

Moreover, what is evident is that very few studies have been done on AFL vocabulary learning and more empirical evidence is needed to shed light on a number of issues including the role of script, orthography, phonology, semantic access, and the impact of morphology and diglossia in both deliberate and incidental AFL learning contexts with L1 English learners.

## Chapter 4: Multi-modal Arabic Word Learning in Novice L1 English Speakers

### Abstract

Word learning studies have shown that individuals can acquire foreign language word knowledge in relatively few exposures to multi-modal stimuli (Bisson et al., 2013, 2014a, 2014b). In the present study, native speakers of British English with no prior knowledge of, or exposure to, Arabic participated in an explicit learning task in which they were exposed to 40 matched and 40 mismatched Arabic vocabulary items accompanied by images and audio. Learning took place over three blocks and results from the initial learning phase indicate participants successfully created form-meaning links. They returned five to seven days later to take part in a second session, this time responding to the same set of 40 “old” words intermixed with a “new” set of 40 words over two blocks. Participants were significantly more accurate and faster in their response times to the “old” words and accuracy scores were only marginally different from those recorded in the final block of the initial learning session. Findings suggest adult learners can acquire Arabic word knowledge after only six exposures in an explicit multi-modal learning task, and that this knowledge is still available to them when testing occurs a week later. At the item level, results were consistent with previous research on processing speed, shorter words containing fewer syllables were associated with faster response times. Additionally, the number of Arabic only phonemes in a word, put differently the number of phonemes that are native to Arabic but do not exist in English, was identified as a difficulty driver. Words with no Arabic only phonemes were associated with higher accuracy rates, and words with two or more Arabic only phonemes, with lower accuracy. Results are explored in terms of current thinking about Arabic vocabulary learning.

**Introduction**

Word learning in Arabic as a Foreign Language (AFL) is often referred to by researchers, teachers and students in terms of its difficulty (Anderson & Suleiman, 2009; Brosh, 2015; Hansen, 2008, 2010; Kuntz, 2000; Mohamed, 2016; Moser, 2013; Ricks, 2015; Ryding, 1991; Young, 2011). In fact, according to Khoury (2008) "…Arabic vocabulary acquisition is argued to be the most difficult aspect of learning Arabic," (p. 2).

As discussed in previous chapters, there are a number of reasons why this might be the case: an unfamiliar script with a deep orthography (i.e., not a one-to-one mapping of graphemes and phonemes), a phonemic inventory with unfamiliar sounds, a lack of cognates, and a different lexical structure (i.e., Arabic words are often formed by the pairing of a root with an inflected form). Yet many of the claims in the literature are made based on teacher observations vs. empirical research (Khoury, 2008). And while vocabulary acquisition has been extensively investigated in the Second Language Acquisition Literature, it is not always possible or appropriate to generalize the findings of studies which use English, Spanish, French and German to non-Roman script and less commonly taught languages like Arabic (Golonka et al., 2015).

At its most basic level, word learning entails the association of a word's form, auditory and/or written, with its meaning (Nation, 2001). Recent approaches to multi-modal learning have shown that foreign language vocabulary can be acquired by novice learners with relatively few exposures when target items are presented with images and audio (Bisson et al., 2013). Studies have shown this approach can work for L1 English speakers learning Welsh (Bisson et al., 2013, 2014a, 2014b), thus the primary goal of the present study is to evaluate the use of similar multi-modal approaches in Arabic.

**Models of Working Memory and Theories of Processing**

To better understand how the auditory and visual characteristics of multi-modal input affect word learning and retention, it is useful to first briefly consider the ways in which such information is perceived, processed, and stored in the brain, as well as the cognitive structures involved in this processing.

Atkinson and Shiffrin's (1968) original multi-store model of human memory divided it into three distinct areas including sensory memory, which was responsible for capturing what an individual sees and hears, short-term memory that could further attend to input perceived by the sensory store, and long-term memory. Craik and Lockhart (1972) contributed to our understanding of the impact of depth of processing, suggesting that greater elaboration and association of meaning could facilitate transfer from short to long-term memory stores, and Baddeley and Hitch (1974) proposed a more detailed system for short-term memory, entitled working memory. They envisioned working memory was made up of a central executive function that controlled attention, and which has since been updated to account for the impact of episodic memory on information transfer (Baddeley, 2001), and to include two secondary systems.

The first of these secondary systems was the phonological loop, which could trap auditory information and rehearse this input to prevent decay, facilitating transfer to long-term memory stores. Papagno, Valentine, and Baddeley (1991) demonstrated the role played by the phonological loop in foreign language word learning in a study of Italian first language (L1) adults tasked with learning Italian-Russian L2 word pairs. The rate of acquisition of the Italian-Russian words presented in written and auditory format was disrupted by articulatory suppression, which prevented the phonological loop from assisting in learning. In contrast, familiar L1 Italian-Italian item pairs were not affected.

The second system was the visuo-spatial sketchpad, now called a scratchpad, which could be used to save pictorial information including images and the written form of a word. In a study of the influence of working memory on vocabulary acquisition from reading, Elsayyad, Evaratt, Mortimore, and Haynes (2017) found it plays a significant role when reading conditions in Arabic are un-vowelized (i.e. deep orthography). Of note, their study involved first language (L1) Arabic readers in the 6-11th grades, not AFL learners who may find working memory is considerably more taxed when reading un-vowelized text in Arabic.

*Using Images in Word Learning*

In the Dual-Coding Theory, Paivio (1971) proposed that when we learn new material we split information between a visual "imagen" based and verbal "logogen" based coding system. The information is saved in different stores and can be connected in an associative, referential, or sensory manner. The idea of separate encoding for pictorial and verbal information is of particular interest to multi-modal learning paradigms as it supports the view that when images and words are learned together it creates a stronger memory that facilitates recall.

Paivio and Csapo (1973) also looked at the superiority of using images over verbal prompts in a seminal study focused on understanding what gives images their advantage in recall testing. In the first of several experiments, the researchers presented 72 concrete nouns, 72 abstract nouns and 72 line-drawn images labeled by the concrete nouns. N=142 participants were given an incidental learning task in which they were briefly shown a randomized list of stimuli with each prompt appearing for 1/16th of a second. They were then given five seconds to write down the word they saw or the label of the picture. Recall was significantly greater for pictures, followed by concrete words and finally abstract words.

In a second experiment, Paivio and Csapo (1973) used a similar procedure but had participants either write a word down or draw an image during the orientation task, which they suggested encouraged the creation of mental images to help encode the words. Stimuli were reduced to only concrete nouns and here the researchers found the superiority of the picture presentation condition disappeared, suggesting mental imagery also assists with word-encoding and aids memory and recall. In their fifth experiment, they created repetition conditions for their stimuli, which followed either a picture-picture, picture-word, word-picture or word-word pattern. They found higher recall for pictures than words when items were presented once and higher recall for the picture-picture and picture-word conditions compared to the stimuli repetitions for word-word.

In looking at more recent research, Curran and Doyle (2011) explored the picture superiority effect on recognition memory. Participants studied lists of words and pictures or just pictures and then read or named them out loud. They were given a recognition test that included new items that had not been studied and ERP measures indicated a stronger old/new effect for pictures. Hockley and Bancroft (2011) also found support for the picture superiority effect in testing participants who had formed associations between concrete word pairs and line-drawn image pairs. Takashima, Bakker, van Hell, Janzen and McQueen (2014) investigated fMRI results when novel words were learned with a phonological form alone or a phonological form paired with an image. They found that when words were learned with pictures the episodic memory system was more involved and retrieval was greater in a 24-hour delayed measure.

## Multi-modal Learning

Multi-modal learning is defined as the presentation of verbal stimuli, in either a written or auditory format, accompanied by pictorial stimuli (Bisson et al., 2014b). The latter may be a still picture, such as a photograph or hand-drawn sketch, or a dynamic image, in the case of video animation (Erce, 2008).

### *Incidental Learning Studies*

Bisson, Van Heuven, Conklin and Tunney (2013) exposed participants to multi-modal stimuli in a letter search task followed by an explicit word learning phase. They presented 40 Welsh concrete nouns paired with line-drawn images and audio to L1 English participants who had no prior knowledge of the target language and were not aware of the learning task that followed. In the deliberate learning task participants heard the word and had to indicate whether or not the English translation equivalent on the screen was a match. They were given words to which they had already been exposed, and new words. Results showed participants were significantly more accurate on previously encountered words. They also had slower response times to the old words compared to a control group. These results suggest they were engaging knowledge gained from the multi-modal exposure to complete the task.

In a follow-up study, Bisson et al. (2014a) used a similar design to focus on the amount of multi-modal exposures needed to produce an incidental learning effect in N=78 native English speakers with no prior experience learning Welsh. Words in the incidental letter-search task were presented for two, four, six or eight repetitions. While Bisson et al. (2014a) were not able to obtain a constant effect per exposure, they did observe that even in old words which participants had only encountered two times in the initial phase, performance was significantly better than for new words. This suggests initial form-meaning links can be made in as little as two exposures to multi-modal input.

In a third study Bisson et al. (2014b) used eye-tracking to look at the relationship between how learners direct their attention in an incidental multi-modal task and their performance on immediate and delayed recall and recognition tests that followed an explicit learning phase. N=66 L1 English participants with no previous exposure to Welsh either heard a word, heard a word and saw its meaning in English, or heard a word, saw a translation and were shown an image depicting its meaning. In the direct learning task participants were required to learn a set of 39 previously encountered words and a set of 39 "new" words using translation equivalents. Results from the free recall and recognition tests showed the incidental learning groups outperformed a control group and were significantly more accurate on previously encountered words. Post-test results also demonstrated an advantage for the group that received an image, with more words recalled one week later and time spent looking at the image in the incidental phase serving as a predictor for the delayed post-test results.

All three of these studies rely on a savings paradigm based on the Ebbinghaus Forgetting Curve (1913) to demonstrate initial learning gains from the incidental phase. Ebbinghaus's Forgetting Curve (1913) explains information decay as a function of exposure repetitions and time intervals. It assumes exposure to a stimulus leaves a memory trace. The memory trace may not be strong enough to facilitate target language word-production on a recall test, nonetheless some word-knowledge exists in memory. By comparing participant performance on new words to old words which had been previously encountered, the difference in the knowledge gained during the incidental phase can be shown as a performance advantage for old words. Bisson et al. (2013) selected the savings paradigm approach because they viewed it as a more sensitive measure of learning gains, but it does leave a degree of uncertainty as to the nature and extent of the word-knowledge that was acquired by participants in the initial learning stage.

*Deliberate Learning Studies*

Researchers have also employed multi-modal approaches in deliberate learning studies in which participants were aware of the learning task. Nassaji (2004) looked at the influence of stimulus modality on name-referent learning with 79 L1 Farsi L2 English university students. Participants were assumed to be of an upper-intermediate level in English and were split into groups to learn name-referent associations with an image and either audio, visual or both audio and visual target word presentations. Results indicated a significant effect of treatment condition with the group who received an image, audio and visual word presentation outperforming the other two treatment groups on recognition and recall measures.

Nassaji (2004) interprets name-referent results as providing support for multi-modal vocabulary presentation when a written form of the target language word is accompanied by audio and visual prompts. Yet the verbal stimuli used were written in all capital letters in Roman script. Thus, the task was more akin to learning an L3, which shared the same script as the participants' L2 of English. Moreover, the pseudowords themselves were mono-syllabic consonant-vowel-consonant letter combinations such as MEV, TEM and TEV made up exclusively of phonemes that existed in both the participants' L1 and L2. It is thus difficult to equate the challenges faced by participants in this learning task to the real world conditions a novice L1 English learner experiences studying Arabic words. There were also relatively fewer items in the stimulus set as compared to previously reviewed studies, namely eight items per block. This may have presented less of a challenge, particularly if participants were relatively experienced language learners.

Zarei and Khazaie (2011) ran a study in which they looked at multi-modal learning using three presentation conditions for participants with different verbal and visual short-term memory abilities. N=161 L1 Farsi L2 English students from a language institute in Iran completed a series of short-term memory and proficiency tests and were then sorted into different verbal and visual ability groups. Each participant was shown a set of stimuli in three different presentation conditions: 1) translation equivalents plus audio and part of speech data, 2) translation equivalents, audio, part of speech data and the word used in a sentence, 3) everything included in the second condition plus an image. Results showed optimal performance was achieved when presentation conditions matched participants' short-term memory abilities. For example, the advantage for the image condition could not be found across all groups; the low-visual and low-verbal ability group performed significantly better on the presentation condition which did not include an image.

While Zarei and Khazaie (2011) forces us to consider the interplay of multi-modal presentations with individual differences in learners' cognitive abilities, the study also presents a very different kind of learning paradigm. Participants had two minutes to learn a word with only one exposure. This is enough of an interval to utilize a range of vocabulary learning strategies that may have affected performance and explained the effect of individual differences on results (Macaro, 2001).

In addition, no discussion of the type of words that were taught is mentioned. The verb "dig" is provided as a sample stimulus item, which suggests the stimuli may have included different parts of speech. A number of researchers have found differences in the way nouns and verbs are acquired and stored in the mental lexicon (Davidoff & Masterson, 1996; Druks, 2002). More specifically, in their study of recall performance comparing rote and keyword learning groups, Ellis and Beaton (1993) found nouns were easier for participants to recall when learned with images. The two studies undertaken in this thesis employed only concrete nouns as stimulus items.

### *Multi-modal in a Classroom Context*

Looking at the research on multi-modal vocabulary presentation in a classroom context is useful because it allows us to consider future applications for this learning approach outside of the language lab. Cárcamo, Cartes, Velasquez, and Larenas (2015) measured the impact of using multi-modal approaches to vocabulary study with $8^{th}$ to $11^{th}$ graders in public schools in Chile. Over the course of five learning sessions 145 students were taught 30 English words using multi-modal input from a slide projector followed by tasks that involved further elaboration through verbal production and additional image pairing. Stimuli were both nouns and adjectives and participants completed a pre and post-test. Significant learning gains were observed across all sessions which the researchers attributed to the benefits of the multi-modal input.

Erce (2008) explored the impact of dynamic vs. still images and cueing vs. the absence of cueing in multi-modal learning, delivered a vocabulary intervention to N=199 L1 English speakers enrolled in 22 sections of Elementary Spanish at an American university. The researcher divided participants into four treatment groups including a static and cueing, static and no cueing, dynamic and cueing, and dynamic and no cueing group. The static group saw line-drawn images, the dynamic group was played short animations, and the cueing consisted of a red flashing highlighted presentation of the target word. The intervention took place over four days and was followed by an immediate post-test and a delayed post-test at 36 days. Results revealed no significant differences between the dynamic and static image conditions on immediate and delayed recognition tests, but the static group outperformed the dynamic group on delayed recall tests.

*Multi-modal AFL studies*

Multi-modal studies in the AFL research have primarily been concerned with evaluating phonemic contrasts as opposed to word learning (Han et al., 2017; Han & Oh, 2018; Showalter & Hayes-Harb, 2015).

Showalter and Hayes-Harb (2015) presented a series of 12 Arabic non-words to N=30 L1 English speakers with no prior experience with Arabic. Stimuli consisted of six minimal pairs that began with ك /k/ or ق /q/, a phoneme contrast English speakers studying Arabic traditionally struggle to differentiate. Pseudowords were created by the researchers and recorded by native speakers, then randomly paired with line-drawn images of familiar items. The initial learning phase consisted of stimuli presentation with audio, images, and a written form in the Arabic script for a total of eight repetitions per item. In the second stage of the experiment, researchers tested to see if participants had created form meaning links by presenting the auditory form of a word alongside a picture. The mean scores for the Arabic script group M=.85 and control group M= .86 on matched items suggest participants learned to associate form with meaning. This is not surprising given the high number of exposures to each word.

Nonetheless, it's difficult to relate these results to an AFL word learning context given the narrow range of the taught stimulus set. As in Nassaji (2004), Showalter and Hayes-Harb (2015) constructed pseudowords in a consonant-long-vowel-consonant configuration instead of using authentic AFL vocabulary items. Moreover, the phonological form of the words was highly restricted given all eight items began with two phonemes that are commonly confused by AFL learners. According to the phonological similarity effect (Baddeley et al., 1984; Schweickert et al., 1990), less accurate recall performance can be expected when participants are listening to lists of words that sound alike. Notably, the group that received the Arabic script did not have an advantage over the control group. We will return to discussion of this study in the following chapter.

Han, Jeon and Oh (2017) employed a multi-modal AFL stimulus presentation in their study concerned with how phonetic categorization is acquired in early learning. Participants were N=28 Korean L1 native speakers L2 English speakers with no prior experience or exposure to AFL. As in Showalter and Hayes-Harb (2015), the stimuli set was comprised of phoneme contrasts, including the ك /k/ and ق /q/ distinction, and additional pairs where one or both members of the contrast existed in Korean. However, in this study researchers selected 36 authentic mono and bi-syllabic MSA nouns to serve as items. These were matched to line-drawn images (Snodgrass & Vanderwart, 1980) and words were recorded by two different Arabic native speakers.

Stimuli were shown twice in randomized order in an initial session and participants were asked to memorize the association and move on when they were ready. In a second-stage participants heard a word and saw two image options on-screen. They had to choose between them and were provided feedback on the correct answer. They were then asked to shadow the word by speaking it aloud. In reviewing this sequence of tasks we can see that Han et al. (2017) employed procedures that encouraged more elaborate processing of the words by participants (Liu & Todd, 2016). It is thus not surprising that by day three participants had already reached a ceiling on a multiple-choice measure, with scores of .99, .98 and .99 across all items. The results of Han et al. (2017) emphasize the importance of being aware of the impact of exposure repetitions and task engagement on word learning results.

Han and Oh (2018) ran a second study with a slightly different design in which they again used multi-modal stimuli to teach minimal pairs of Arabic non-words. Stimuli were 30 mono and bi-syllabic Arabic pseudowords which were transcribed in one way for one treatment group and in a different way for another, based on the target phoneme contrast the researchers were testing. A bi-modal control group was also included who received only images and the auditory form of the word during the initial learning stage. Participants were N=47 Korean L1 English L2 speakers with no prior experience studying AFL. All words were randomly paired with line-drawn images (Snodgrass & Vanderwart, 1980) and audio was recorded by Arabic native speakers.

The learning presentation resembled Han et al. (2017) but participants were additionally told to pay attention to spelling. In the second phase only image options and audio were played to test the participants' ability to map spoken form to meaning and on day four researchers asked participants recall the spelling of the target words by writing them down by hand when cued with an image. Results showed no significant differences between treatment groups on form-meaning test performance and a lexical decision task. However, in interpreting spelling results, the auditory only group significantly underperformed with 62.7 accuracy compared to 99.3 and 96.7 for the two spelling treatment groups. This is understandable as the auditory input only group were left to devise grapheme-phoneme correspondences on their own.

What's unclear in Han and Oh (2018) is whether the two spelling treatment groups were shown words written in Korean letters, and whether they had to write these words in Korean or Roman script for the spelling test. Researchers specify that the auditory only group was instructed to use their L2 English to write the words using Roman letters and an appendix containing stimulus items seems to suggest the same was true for the treatment groups. This may have conflated results given stimuli were chosen to exploit L1 Korean sound contrasts which did or did not exist in MSA, but were encoded graphically using L2 English letters.

In reviewing AFL studies that involve multi-modal learning, we can see different choices made in regards to stimuli sets. Two studies used non-words, while the other worked with authentic AFL concrete nouns. More importantly, all of the stimuli sets in the AFL research examined here contained a more phonemically restricted range of words than a learner would encounter studying a textbook-based vocabulary list.

*Summary of Research Reviewed*

Overall, the studies reviewed have demonstrated the efficacy of multi-modal approaches in vocabulary learning – and the need for more word-learning focused research in an AFL learning context. They also make clear that aspects of study design can be manipulated to produce various effects. For example, larger stimulus presentation windows seem to invite participants to engage in personal vocabulary learning strategies, which can mask the impact of multi-modal stimulus presentation. This is one reason why the present research worked with a narrow stimulus presentation window and directly emphasized the automaticity of the task to participants. The number of to-be-learned items, diversity of word forms within the stimulus set, and number of exposures can also affect results and for this reason all three variables were carefully controlled in the present study.

## Present Study and Research Aims

The present study focuses on the use of multi-modal input in early AFL learning with novice L1 English speakers. AFL vocabulary learning is consistently referred to as "difficult" in the literature, in the classroom, and even in the introduction to many Arabic textbooks (Belnap, 2006; Cicerchia, 2010). This creates an aura of impossibility around the learning of Arabic for English speakers, one that is pervasive and proliferated by AFL students and teachers alike (Kuntz, 2000). Thus, a secondary aim of this research is to debunk some of the myths around the impossibility of learning vocabulary in Arabic, and show that with appropriate approaches, early learning gains can be made, even in adult learners with no prior experience with the language.

Research questions for the study were the following:

1. Does AFL vocabulary learning occur in L1 English learners who participate in a multi-modal learning task?
2. How do the properties of the stimuli (length, number of syllables, number of Arabic only phonemes, frequency, concreteness, image complexity) explain variance in accuracy scores and response times?

## Method

### *Participants*

Thirty-two Native English speakers recruited from the University of Nottingham's UK Campus participated in this research. They included undergraduate and graduate level students and university workers who were made aware of the Arabic word-learning task through email and poster advertisements. The email and poster advertisements and the initial responses from the researcher to individuals who expressed an interest in taking part, emphasized that they should have no prior knowledge of or exposure to the Arabic language in order to participate in the study. Participants read a project information sheet and signed an ethics form before beginning the experiment. They were compensated financially for their time. All recruitment and participation procedures were approved prior to the start of the research by the University of Nottingham's School of English Ethics Committee.

Seven Males, 22 Females and one Individual who chose not to disclose their gender took part in the study. Their mean age was M= 23.4 years with a range of 18-35. Twenty-five of the participants were pursuing or had received their bachelor's degree, three were pursuing or had received a master's, two were in a doctoral program and two chose not to disclose their highest level of education. Before beginning the experiment, participants were again asked if they had previous exposure to the Arabic language or had spent extensive time in an Arabic speaking country. Only participants who responded "no" were included in the study. A further question confirming they had no prior exposure to the Arabic language appeared on a language background questionnaire taken by participants before beginning the study.

Thirty-one of the 32 participants had studied a language other than English before, with more than half reporting intermediate or advanced proficiency. The most frequently studied languages included French, Spanish, German, and Mandarin. Participants also reported experience with Italian, Korean, Punjabi, Dutch, Welsh and Malay. This data was collected using a language background questionnaire with the goal of investigating whether previous experience learning non-European languages, learning different script languages, and/or achieving higher fluency levels in second or foreign language study, would impact participants' performance on the Arabic word learning task.

*Design*

The researcher adopted a version of Bisson et al.'s (2013, 2014a) savings paradigm approach in which participants were initially presented with a set of words to learn and then returned for a second session in which previously encountered words were presented alongside new words. The number of stimulus items and blocks were based on Bisson et al.'s previous research and care was taken to select a number of exposure repetitions that would facilitate word learning without causing participants to reach an early ceiling. It was also important to choose a testing approach that would be sensitive enough to show learning gains.

Stimuli were counterbalanced and semi-randomly assigned to two lists; half of the participants only saw List 1 in the initial session, and half only saw List 2; both groups saw Lists 1 and 2 in the final session. Word type (old and new) and match type (match and mismatch presentation) were the within-subject factors.

*Materials*

Stimuli included 80 concrete Arabic nouns (Appendix 4.1). Each item presentation consisted of a black and white line-drawn image to represent meaning, a transliterated written word form displayed below the image, and a native-speaker audio recording of a spoken word form that matched the written form. Participants either saw an image that correctly represented the word's meaning or an image that corresponded to another of the taught stimulus items. Additional details on how mismatch images were assigned is provided at the end of this section.

The pictorial stimuli used were normed by Snodgrass and Vanderwart (1980) and Cycowicz, Friedman, Rothstein, and Snodgrass (1997). The decision to use these images was based on their widespread use in past research.

Attention was paid to the nature of the vocabulary items taught, with the goal of providing a degree of authenticity to the learning task by ensuring the diversity of the stimulus set as a whole. In selecting the stimuli, the researcher began by reviewing a set of 300 nouns representing common household objects, animals, and food. This genre was selected based on its ecological validity, as it was believed the learning challenge would be more akin to early word learning tasks undertaken by beginner language learners. Working from an initially larger set of words and reducing it down in stages also helped to preserve a measure of variability in the final stimulus selection.

All of the items in the original list were concrete nouns with high imageability ratings (Snodgrass & Vanderwart, 1980). Given the diglossic situation affecting Arabic and the prevalence of dialects which can range greatly in their use of alternative labels for everyday items – see previous discussion or Ryding (1991), the researcher enlisted a panel of five native speakers from different regions of the Arabic speaking world to screen the initial set of Modern Standard Arabic translation equivalents. Words with labels that could not be agreed upon were removed from consideration.

The remaining words were then reviewed for cognate status. As previously discussed, cognates are words which share a similar form and meaning in two languages. The decision to exclude cognates from the stimulus list was made based on previous research which has shown cognate status can affect learnability (Lotto & de Groot, 1998). Arabic and English have a relatively fewer cognates than European languages; however, several food, animal and technology related loan words had overlapping form and meaning and were therefore removed. The list was also reviewed for cognates between Arabic and Spanish and Arabic and French, given these languages were predicted to be familiar to some of the participants. Additionally, two items were removed from the final analysis due to unforeseen cognate status – this will be discussed in greater detail in the following section.

Finally, any words that were culturally incongruent for Arabic, for example "baseball bat," were removed, and those which may have elicited negative memories and/or emotions from participants, such as "gun." This was based on research that has shown emotionally charged items can elicit different amounts of attentional resources (Schupp et al., 2007). It is also best practice in language testing to avoid the use of inflammatory or upsetting language whenever possible ("ECCE Guidelines for item writers," 2014).

The Arabic words were then transliterated by the researcher according to an adapted form of the Qalam transliteration system – see previous discussion on transliteration and transcription issues in Arabic, and the final word list was checked by a native speaker for errors. Transliterations ranged in length from three to ten letters.

Audio recordings of the target list were made by a male native Arabic speaker from Saudi Arabia. While the voice actor had previously recorded language learning material for an AFL L1 English audience, instructions were given to read the words as naturally as possible and to provide three recordings of each one so the best one could be selected by a panel of two native speaker raters who came from different L1 Arabic dialect backgrounds. Recordings were made in a professional sound studio using a Dell Desktop PC and AKG C 2000 B microphone and the final edited WAV files were produced using Adobe Audition. All recordings were reviewed by a panel of native speaker raters and several rerecords were provided based on their recommendations.

The transliterations and audio were used to identify Arabic-only phonemes in the words, with Arabic-only phonemes (AoPs) defined as sounds which do not exist in the participants' first language of English. In the final set there were 36 words which contained 0 AoPs, 34 words which contained 1 AoP, and seven which contained 2+ AoPs, including one word with 3 AoPs.

English word frequency ranged from 1.97 to 6.43 in the SUBTLEX-UK Frequency List (van Heuven et al., 2014) and words were split into five frequency bands. Stimuli were semi-randomly assigned to two lists, which were balanced for word length (number of letters and syllables), Arabic only phonemes, and semantic overlap. In the final stimulus set all but five of the total possible initial-onset-phonemes in Arabic were represented.

The researcher created 80 match pairs and 320 mismatch pairs. Mismatch pairs were assigned randomly within lists and did not repeat in Blocks 1, 2, or 3, or in Session 2. Care was taken to ensure that mismatches were not phonetically or semantically related to the match word. The image either matched or mismatched the accompanying word.

*Procedure*

In the initial session, participants were provided with a study overview document to read, an ethics form to read and sign, and a participant background sheet to fill out. It was then explained to them that they would be learning a set of 40 Arabic words in three rounds. They were told they would see an image, hear an audio, and be presented with a word on screen. Their task was to decide if they felt the audio and written word matched the image presented. They could indicate their responses using the “yes” and “no” keys on the keyboard.

The researcher read out a standard script to each participant before they began which emphasized the automaticity of the task. It was explained that they would have to guess in the initial round and then would be provided with feedback which could be used to learn the words. Feedback was in the form of the word CORRECT written in green or INCORRECT written in red. All instructions and feedback were provided in English. Once feedback had been provided, the next item automatically appeared on the screen, meaning there was no opportunity for some participants to study this feedback longer than others. Participants were encouraged to respond as quickly and accurately as possible to the stimuli. They were told they would see their percentage correct score at the end of each round. These instructions were provided verbally and on-screen before the participants began the experiment.

Testing was done individually by the researcher in a quiet and distraction free environment in the University of Nottingham's Psycholinguistics Lab. The stimulus material was displayed using E-Prime version 1.0 on a PC computer. Participants used headphones and the researcher checked with them to ensure volume was sufficient before they began the first block. Participants were asked to turn their phones off and advised that they could take a break at the end of each block if needed.

The first session consisted of three blocks. Each word appeared once in a match presentation and once in a mismatch in each block for a total of six presentations in Session 1. Presentation order was randomized for each participant. Percentage correct was shown at the end of each block, and the initial three blocks took between 15 and 25 minutes to complete.

Session 2 took place five to seven days later. Participants were advised they would be doing the same activity as in Session 1, except material would be presented over the course of two blocks instead of three. Instructions were again provided on-screen, and participants were shown each of the original words presented in Session 1, in both a match and a mismatch presentation, as well as a set of new words in a match and a mismatch presentation. Stimulus order was randomized within blocks for all participants. All participants completed Session 2 in less than 15 minutes.

## Results

### *Preliminary Analyses*

Accuracy and response time data was analyzed by participant and by word. By participant, average accuracy scores and response times for old and new words in the match condition were calculated and screened for outliers. In looking at outliers, particular attention was paid to three participants who completed their second session outside of the five to seven day window. No outliers were identified, apart from Participant 19 who had exceptionally long response times. This may have been due to her background as a bilingual English L1 Punjabi L2 speaker. As accuracy rates for this participant were not outliers, it was decided to keep her in the analysis. Average response times in Session 1 ranged from 1046.79 to 2668.43 ms, $M$=1728.29 $SD$=430.54, indicating some participants were generally faster to respond to items than others. However, no significant correlation was found between average response times and accuracy, thus taking more or less time to respond to items did not necessarily impact task performance.

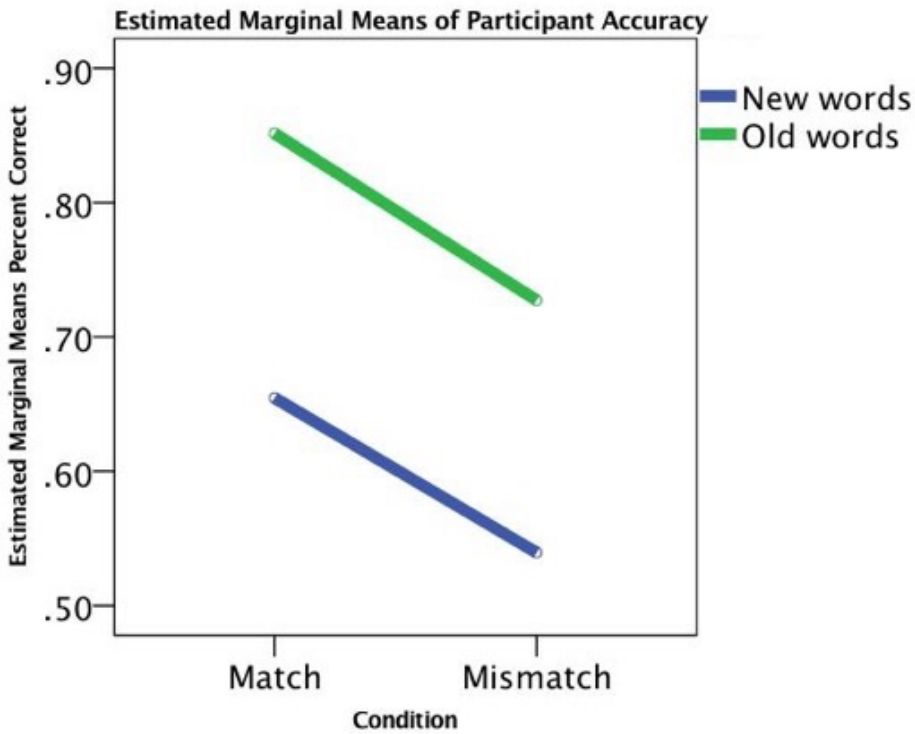
A further analysis of the relationship between participant accuracy in Sessions 1 and 2 and days between sessions was undertaken to ensure that data from participants whose sessions were two, eight and 11 days apart should not be removed. No significant effects were found when their data was included, thus it was decided not to remove them from the final analyses. The total participant counts were Group 1, N=17 and Group 2, N=15.

A visual inspection for outliers in word data revealed "miTraqa" (meaning "hammer") had low accuracy in the match condition; however, as it was still within two standard deviations of the mean, it was included in the final analyses. Following the completion of Session 1, several participants mentioned having identified a word they found easy to learn. Upon closer inspection, it was decided that the word "istable" was a cognate with "stable" in English and therefore easily associated with "farm." This word was consequently removed from the analysis for all participants. The word "Tufah" or "apple" was also removed from the data set because "istable" was the primary distractor used in its mismatch conditions. Both of these words came from List 1, thus the total final item counts were List 1=38 and List 2=40. Additionally, one participant identified a Hebrew Arabic cognate "yad" for "hand" and reported this to the researcher. This item was removed from the analysis for this particular participant. Total data loss was 2%.

Independent sample *t*-tests revealed no significant differences between groups were found (*p*'s > .05). However, in looking at lists, there was a significant difference in accuracy scores for old words. List 1's mean accuracy in a match condition was *M*=.82 *SD*=.11 and List 2's *M*=.88 *SD*=.08. Independent sample *t*-tests showed $t(76)=-2.56$ $p<.05$ with a moderate effect size eta squared .0791. Thus 7.91% of the variance in accuracy can be explained by list. The researcher hypothesized this may have been due to data loss in List 1 where two items which participants generally performed well on were removed because of unforeseen cognate status. There were no significant differences in average response times between the two lists (*p*'s > .05).

**Figure 4.1**

*Estimated Marginal Means of Participant Accuracy for Old and New Words in a Match and Mismatch Condition*



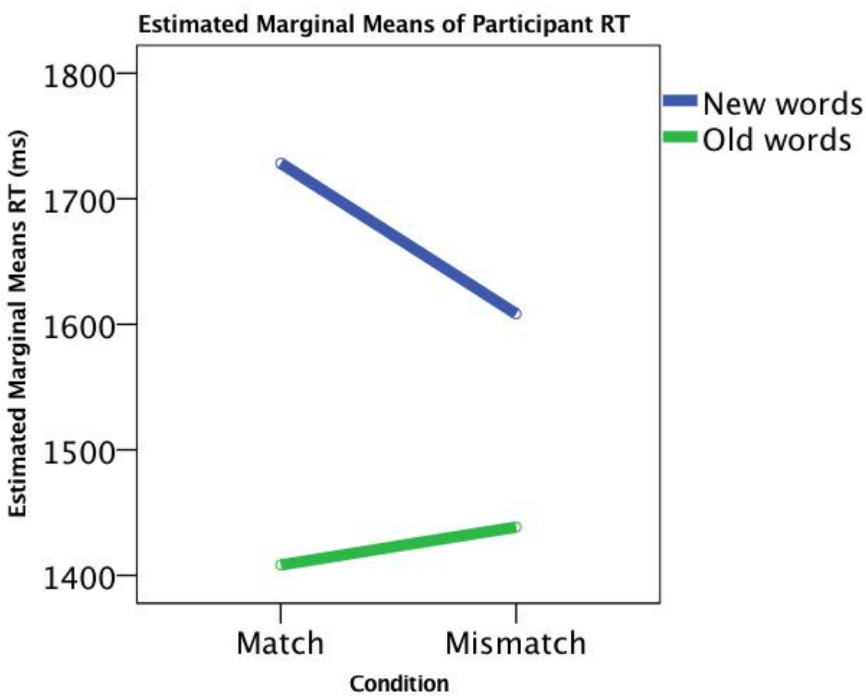

*Accuracy*

A Two-Way Analysis of Variance (ANOVA) was carried out to check for main effects of Word Type (old/new) and Presentation Condition (match/mismatch) in measures of accuracy by participant and by word. As shown in Figure 4.1, results revealed a significant effect for accuracy on old vs. new words $F_1(1,30)=149.88$, $p<.001$, Partial eta squared =.69, $F_2(1,76)=149.33$, $p<.001$, Partial eta squared =.66 with large effect sizes. There was also a significant effect for match vs. mismatch presentation $F_1(1,30)=15.03$, $p<.001$, Partial eta squared=.33, $F_2(1,76)=88.76$, $p<.001$, Partial eta squared=.54. Participants were significantly more accurate on old words they had seen in

Session 1 than new words, and they were more accurate in a match presentation for both old and new words. The interaction between old/new and match/mismatch was not significant: $F_1(1,30)=.05$, $p=.82$, $F_2(1,76)=.10$, $p=.75$.

**Figure 4.2**

*Estimated Marginal Means of Participant Response Times for Old and New Words in a Match and Mismatch Condition*

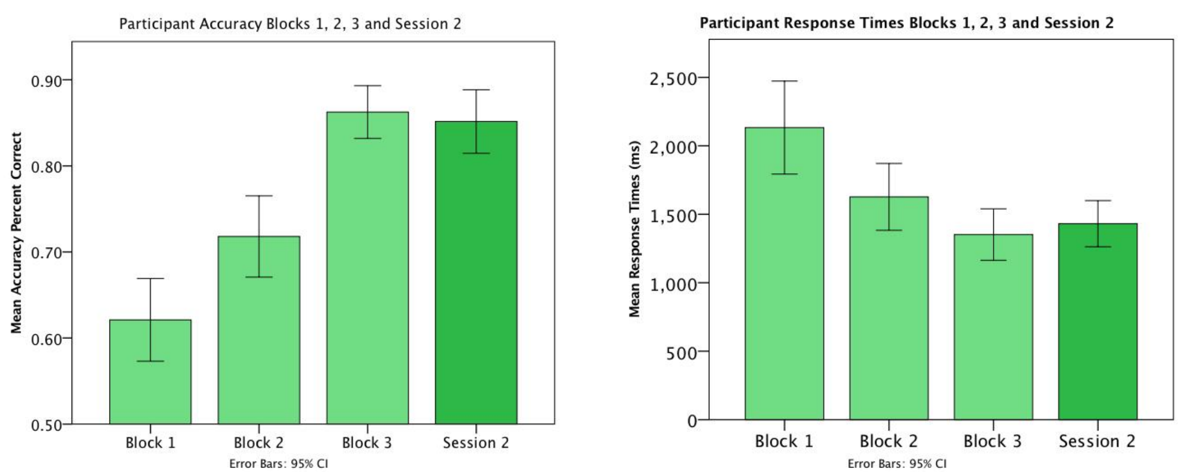



*Response Times*

A Two-Way Analysis of Variance (ANOVA) was carried out to check for main effects of Word Type (old/new) and Presentation Condition (match/mismatch) in measures of response times by participant and by word. As shown in Figure 4.2, results revealed a significant main effect of old vs. new words $F_1(1,30)=67.98$, $p<.001$, Partial eta squared =.69, $F_2(1,76)=131.00$, $p<.001$, Partial eta squared =.63, with large effect sizes. There was also a significant interaction between Word Type and Condition $F_1(1,30)=16.99$, $p<.001$, Partial eta squared=.36, $F_2(1,76)=6.87$, $p<.05$, Partial eta squared=.08. Participants were significantly faster with old words than new words and the effect of a match/mismatch presentation was greater in new words than old words and in the opposite direction. In other words, participants got faster in a mismatch presentation for new words, whereas the opposite happened for words they had seen in Session 1. Response times in match/mismatch conditions were not significant $F_1$ $(1,30)=1.08$, $p=.30$, $F_2$ $(1,76)=.019$, $p=.89$.

**Figure 4.3**

*Mean Participant Accuracy and Response Times for Old Words Blocks 1,2,3 and Session 2*




*Note.* Error bars show standard errors: Confidence interval 95%.

*Results Across Blocks*

As can be seen in Figure 4.3, further analyses of the learning data for old words by block revealed participants generally performed at chance level in terms of accuracy in Block 1 of Session 1 and improved steadily with accuracy increasing and response times decreasing by Block 3. No significant differences were found between accuracy and response times in Block 3 and Session 2 ($p$'s>.05). This suggests participants still had access to the learned forms when they returned to take part in Session 2 a week later.

*Number of Letters, Syllables, and Arabic-only Phonemes*

In looking more closely at the data on the number of letters in the transliterations and their effect on accuracy and response times, stimuli were divided into three groups: Group 1: 3-5 letters, Group 2: 6-7 letters and Group 3: 8-10 letters. As can be seen in Table 4.1, a one-way between groups ANOVA and a one-way between-groups ANOVA with planned comparisons for Group 1 and Group 3 were performed to check for differences in accuracy between words of different lengths. No significant differences were found ($p$'s>.05). However, accuracy in the mismatch condition for old words did reach significance ($p$<.05).

Response time data for number of letters also reached significance across all conditions old match ($p$<.05), old mismatch ($p$<.01), new match ($p$<.05), and new mismatch ($p$<.01). Post-hoc comparisons using Tukey HSD indicated short words had significantly lower response times than medium and long words, but there was no statistically significant difference between the latter two. This is consistent with research findings on word length and processing times.

In looking at the impact of number of syllables, words were again split into three groups. As seen in Table 4.1, differences in accuracy were only significant in the mismatch condition for old words ($p$<.01). Post-hoc comparisons indicated 1-syllable words were significantly easier than 2 and 3-syllable words ($p$<.05), but again this was only for the old words in a mismatch presentation. The results were quite different in

**Table 4.1**

*One-Way ANOVA Results for Number of Letters, Syllables, and Arabic-only Phonemes*

| | Number of Letters | |
|---|---|---|
| | Accuracy | Response times |
| | *F* | *F* |
| Old Match | 2.20 | 5.94* |
| Old Mismatch | 3.29* | 9.17** |
| New Match | 1.12 | 4.91* |
| New Mismatch | 0.46 | 9.23** |
| | **Number of Syllables** | |
| | Accuracy | Response times |
| | *F* | *F* |
| Old Match | 1.95 | 4.10* |
| Old Mismatch | 6.89** | 7.22** |
| New Match | 0.91 | 3.81* |
| New Mismatch | 0.32 | 9.57*** |
| | **Number of Arabic-only Phonemes** | |
| | Accuracy | Response times |
| | *F* | *F* |
| Old Match | 2.59 | 0.04 |
| Old Mismatch | 0.02 | 0.38 |
| New Match | 1.30 | 2.13 |
| New Mismatch | 7.86** | 0.59 |

*Note.* ANOVA = analysis of variance

*$p < .05$. **$p < .01$. ***$p < .001$.

terms of response times. Post-hoc comparisons indicated 1-syllable words in a match and mismatch had significantly faster response times than 3-syllable words ($p$<.05). The effect was stronger in new words than old. This is in line with research on increased processing time for longer words.

A one-way between-groups ANOVA was run to look at the impact of the number of Arabic only phonemes (AoP) on accuracy and response times – see Table 4.1. Items were divided into three groups depending on the number of AoPs they contained: Group 1: 0 letters, Group 2: 1 letter, Group 3: 2-3 letters. Note that only one word “Hidhaa” had three AoPs. There were 36 words with no AoPs, 34 with one AoP, and seven words with two to three AoPs. No significant differences were found in the response time data. However, a one-way ANOVA with planned comparisons revealed a significant difference in accuracy between old words with no Arabic phonemes and two or more $F(1,75)=4.22$, $p$<.05. It thus appears old words were easier when they had fewer AoPs and harder when they had more.

A standard multiple regression analysis was also performed to ascertain which factor (letters, syllables, AoPs) contributed most to accuracy and response time. For accuracy, the model explained 9% of the variance in accuracy with number of Arabic phonemes making the only significant contribution (beta=-.23) $p$<.05. For response time, the model explained 14.2% of the variance with number of letters making the only significant contribution (beta=.40) $p$<.05.

*Additional Analyses*

No effects for word frequency were found in old words. However, there was a negative correlation between response time and new words in a match condition. Higher frequencies were associated with faster response times $r$=-.32, $n$=73 $p$<.01.

Participant data was also analyzed see if there was a relationship between individual differences among participants for accuracy and response times. Results showed older participants were significantly slower in match conditions for old $r$=0.36, $p$<.05 and new words $r$=.47, $p$<.001, but not less accurate. No effects were found for gender, highest level of education, or number of languages studied. A one-way ANOVA was performed to explore the impact of highest level of foreign language proficiency achieved based on participant self-report data. No statistically significant difference for accuracy on old words in a match condition was found $F$(3,28)=.30, $p$=.82. An independent samples $t$-test was also performed to compare mean accuracy on old match words for participants who had studied a non-European language N= 11 ($M$=.89, $SD$=.10) compared to those who had not N=20 ($M$=.83, $SD$=.10) but no significant differences were found $t$(29)=-1.11,$p$=.28.

**Discussion**

The primary aim of this study was to investigate multi-modal learning as it applies to Arabic vocabulary acquisition in novice L1 English learners. Results clearly show that participants learned the target words. Not only did they learn them in 6 exposures (3 matched, 3 mismatched), but they retained them a week later, performing with the same accuracy and response times in Session 2 as they had at the end of Session

1. Given the word lists in this study were both robust, including upwards of 38 items, and ecologically valid (i.e., the learned words would not be out of place in a basic or beginner level AFL vocabulary lesson), results are encouraging.

In recommendations to researchers and educators on teaching Arabic vocabulary, Young (2011) emphasizes the importance of finding new ways to lower the learning burden for students, suggesting auditory and visual presentation of words increases the chances of meaning recall. Prior to the current study, little AFL specific evidence existed to sustain this claim. Those multi-modal studies that did involve AFL for L1 English speakers and were reviewed in the previous sections, namely Showalter and Hayes-Harb (2015), were concerned mainly with phonetic discrimination tasks, not word learning. As such, they utilized pseudowords and stimulus items that featured a restricted phonemic inventory.

Participants in Showalter and Hayes-Harb (2015) demonstrated the same mean accuracy rates as were seen in the current research, but they were given far more exposure to items (a minimum of eight repetitions) and had a smaller word list to learn (12 items). Bisson et al. (2014a) used a similar amount of to-be-learned items in their study of incidental learning in Welsh and sought to determine a set amount of repetitions, two, four, six or eight, that could demonstrate learning. While their results were not conclusive, the present study showed learning had occurred after 6 repetitions, with only three exposures in a match condition. Again, this is encouraging as language learners often have limited resources in terms of time, thus the more efficient they can be in their vocabulary study, the better. It is also a sign that whether it be Welsh or Arabic, the pairing of images and audio with stimulus items is an effective learning strategy for vocabulary. Of note, it would be of great interest for future research to further investigate differences in multi-modal presentation. For example, how might learning be affected if no written word form is presented.

We have seen in the literature that the more cognitively demanding and robust the learning task, the greater the results in terms of vocabulary gains (Golonka et al., 2015). The two multi-modal AFL studies reviewed here which involved L1 Korean learners (Han et al., 2017; Han & Oh, 2018) both featured increased exposure to the stimulus items and more opportunities for elaboration of learning, including shadowing activities. On the other hand, the participants in the present research took part in a relatively low-effort activity. They sat at a desk in a language lab, watched a screen, listened to audio, and pressed a key, but still achieved the target learning goals.

Language learners may not always be able to engage in highly contextualized and cognitively demanding tasks when it comes to vocabulary. Quite often, they are provided with words in list form as a means to support other more involved comprehension-based activities, such as reading, speaking, listening, or writing. A multi-modal approach similar to the one taken in this study thus caters to students who need a reliable and efficient way to learn Arabic words.

Moreover, in today's online world, app-based learning is becoming the norm. Many people have tried a language learning app on their smartphone, and even in university-level language courses, apps are recommended as appropriate tools to supplement learning. The results demonstrated here thus serve as support for multi-modal, e-learning-based approaches to vocabulary study in AFL. They may also help inform app development in terms of optimal repetition cycles and desirable stimulus components to be shown alongside words.

Nonetheless, despite the fact that this study's results clearly demonstrate learning, it is important to remember no direct measure of recall was required of participants. When a savings-based approach is taken, there is a degree of unknown as to the extent of the lexical form and meaning learning that has been achieved. Additionally, looking at the response time data shows us that older participants did generally take longer in responding to items, meaning they may have approached the learning task slightly differently. For this reason, more research involving receptive and productive tasks and individual differences, including age and motivation to learn, would be of benefit to researchers, teachers, curriculum and app developers.

In response to the second research question raised in this study which pertains to sources of difficulty in Arabic learning, the data suggests the number of letters and syllables a word contains are not necessarily indicators of an increased learning burden, or at least one that negatively impacts accuracy in learning form-meaning associations. What did come up as a significant factor influencing difficulty was the number of Arabic-only phonemes (AoPs) a word contained. A word with one or fewer AoPs, for example "najma" or "dhura" was found to be easier than a word containing two or more, for example, "Diffda'" and "Taa'ira." This is perhaps a finding that will not be surprising to researchers working in AFL given the abundance of studies reviewed here that look at phonemic contrast acquisition involving AoPs. One caveat to the present findings though is the number of two or more AoP stimulus items featured in this study was quite low, seven, in comparison to 36 and 34 for the no AoP and one AoP groups respectively. Thus, it is important to look at the impact of AoPs more carefully in a follow up study.

It should also be restated that the written forms participants learned were in a Romanized script. While casual Arabic study for purposes of travel or business may not require learning the Arabic alphabet and mastering its orthography, most beginner AFL learners will need to learn words using the Arabic script. As such we will investigate the application of a similar multi-modal approach when target words are presented in a modified version of the Arabic script in the study detailed in the following chapter.

On a final note, it may be relevant to call attention to the language background data obtained by participants who took part in the present research. Most participants had experience with foreign language learning which is in line with trends found by Kuntz (2000) and identified as typical of an AFL learner population.

Kuntz's (2000) survey of American AFL learners in study abroad programs found that 95% of the 71 students surveyed had previously studied other languages. In the US, Arabic is generally considered a less commonly taught language and is infrequently offered to students outside of higher education. If AFL study is pursued, it is following several years of high school education which commonly includes tuition in a second language, such as Spanish or French (Belnap, 1987). Learners in the UK are also less likely to be offered Arabic as a Foreign Language before Key Stage 4 (Ramezanzadeh, 2016), and generally have prior second language experience before beginning Arabic course-work at university level. According to Ramezanzadeh (2016), this holds true for heritage learners as well. She found 62.7% of the AFL heritage learners she surveyed identified culturally as Asian, with predominantly Pakistani and Indian ethnicity cited, and other languages, such as Urdu, listed as spoken at home. Thus, the participants who took part in the present research may likely resemble a typical L1 English AFL learner population, which can help support generalizability of the study's findings to AFL learning in an L1 English setting.

## Chapter 5. The Impact of Letter Training on Multi-modal Word Learning in Arabic

### Abstract

There have been mixed findings as to the efficacy of providing novice L1 English Arabic as a Foreign Language (AFL) learners with L2 written forms as they may receive little to no learning benefit from the target language script (Mathieu, 2016; Showalter & Hayes-Harb, 2015). The current study explores the impact of running a letter training ahead of a multi-modal word learning task. Research on letter training (i.e., explicit instruction in grapheme-phoneme correspondences) has shown it can facilitate word recognition and learning for real and pseudo languages (Bitan & Booth, 2012; Bitan & Karni, 2003; Bolger, 2007). To the best of the researcher's knowledge, only one study has employed an Arabic letter training with English native speakers. However, participants in Bishop (1964) were not given AFL vocabulary items, but rather English words written in the Arabic script. Moreover, no measure of word learning was included. The present study compared immediate and delayed word learning by novice learners of Arabic who were taught the grapheme-to-phoneme correspondence for eight Arabic letters, to that of learners who had no letter training. Results showed a significant advantage for a high-performing subset of the letter-training group in immediate and delayed word recognition and recall tasks. The findings support the use of letter training at the introductory level and suggest that novice learners can make use of the Arabic script to support early vocabulary learning.

**Introduction**

The previous chapter of this thesis explored Arabic as a Foreign Language (AFL) vocabulary learning when words are presented with images, audio, and a transliteration. Nonetheless, one of the earliest challenges English native speakers face in AFL is learning the unfamiliar writing system. As Hansen (2008) points out, this includes mastering a set of novel graphemes (i.e. letters) and their corresponding phonemes (i.e. sounds), as well as a new orthographic code (i.e. the set of conventions that govern a writing system).

There are a number of characteristics specific to Arabic which make mastery of the writing system difficult (Hansen, 2010; Mathieu, 2016). There are 28 letters in the Arabic alphabet, however, only 10 have unique forms. The remaining letters can be subdivided into groups where graphemes share the same basic shape and are distinguished by the presence, number and/or location of a dot. Arabic is also a cursive script in which letters can have up to four allographs (i.e., forms) depending on where they fall in a word. In exploring visual word recognition in Arabic, Boudelaa, Noris, Mahfoudhi, and Kinoshita (2019) concluded it is the position-dependent allographic nature of its script that really "sets Arabic apart from other languages," (p. 744).

The Arabic script is made up of connecting and non-connecting letters (i.e., a space does not always delineate a new word); research has shown that changes to inter-word spacing can hinder reading and word recognition (Bassetti, 2009). Furthermore, as opposed to English, Arabic is read right to left. Differences in directionality between a learner's L1 and L2 can cause processing delays in L2 reading (Randall & Meara, 1988). Moreover, given the distinctive morphology of Arabic (Khoury, 2008), Arabic words have more visual neighbors than English words, which can slow word recognition (Boudelaa et al., 2019). Another consideration is the lack of written short vowels. Both Arabic and English are alphabetic languages. However, Arabic has what has been referred to as a "consonantal system" (Cook & Bassetti, 2005) because short vowels are represented by diacritic marks which are often left out of the text. This poses a challenge for learners and makes decoding more difficult, as not all of a word's phonemes are represented.

Studies of bilingual readers in which Arabic is their first language (L1) and Hebrew their second language (L2), have suggested letter processing is actually slower in the participants' L1 because of the graphical complexity of the script (Eviatar et al., 2004; Ibrahim et al., 2002). Thus, we might also expect processing delays for low proficiency L2 Arabic learners who have yet to develop automaticity in decoding (McLaughlin, 1987).

Research with Arabic as a Foreign Language (AFL) learners has indeed shown that the Arabic writing system can affect students at various proficiency levels. Khaldieh (1996) examined word recognition in beginner, intermediate and advanced level L1 English L2 Arabic learners. Arabic target words were presented on a screen and then removed. Next, participants were offered a choice of visual, phonological, and semantic distractors. Beginner and intermediate learners were more likely to choose visual distractors. The researcher attributed this to difficulty in letter recognition caused by the cursive script.

Hansen (2010) in a study of the impact of the Arabic writing system on visual word recognition, compared the performance of beginner, intermediate and advanced proficiency AFL learners to native speakers on a decoding task. While learners took significantly longer to decode passages written in the Arabic script compared to a control group of native speakers, there were no differences in response times for a passage of Arabic pseudowords written in the Latin script. In Brosh's (2015) study of second and third year AFL learners' spelling errors, wrong letter choice accounted for 32% of all mistakes made, suggesting learners encountered difficulties with both phoneme-grapheme correspondence and the Arabic script.

Khaldieh (1996), Hansen (2010), and Brosh's (2015) studies all point to potential challenges caused by the Arabic script for participants who are already enrolled and studying in AFL programs. What then is the situation for complete novices who have no experience with the script or knowledge of Arabic orthography? While some apps and learning programs may take the approach of providing transliterations to beginner learners, most students will be charged with simultaneously learning the alphabet and their first Arabic words. More research is thus needed to understand how we can best facilitate a combined approach.

### Orthographic Input and Phonological Awareness

Research on orthographic input in early second language (L2) acquisition is often concerned with its impact on lexical learning and the development of phonological awareness in the L2 (i.e., awareness of the set of sounds that make up a language and rules concerning how these sounds interact) (for a review see Bassetti, Escudero and Hayes-Harb, (2015).

Two studies have explored the impact of the Arabic script on word learning and the acquisition of novel phonemic contrasts for beginner learners (Mathieu, 2016; Showalter & Hayes-Harb, 2015). As previously discussed, Showalter and Hayes-Harb (2015) carried out a series of experiments with native English speakers who were tested on their ability to learn a set of six minimal pair Arabic pseudowords that began with either the voiceless velar stop /k/ or the voiceless uvular stop /q/. Participants saw written word forms in Arabic, paired with auditory and pictorial stimuli then took part in a passive training phase before completing a criterion test to measure word learning. Results showed no significant difference between participants who received an Arabic script written form and a control group.

Researchers posited that the inability of the Arabic script group to make use of the written input may have been due to the cursive nature of the script. Indeed it has been shown that visual word forms that are difficult to parse can result in low form knowledge and reduced retention in a learning task (McCandliss et al., 1997). Results may also have been related to the difficulty of the phonemic contrast being tested. Lababidi (2016) in a study of the perceptual mappings of Arabic consonants found the Arabic letter ق /q/ was frequently categorized with a high goodness of fit for /k/ by American English speakers. This was true in a number of different vowel contexts and may also have been true for Showalter and Hayes-Harb's participants. Additionally, Escudero (2015) underscores that orthographic input may only help with minimal pair and novel contrast learning when two phonemes are already perceptually easy for participants to discriminate between, which was not the case in Showalter and Hayes-Harb (2015).

It is important to consider factors related to Arabic orthography as well (e.g. Arabic is written right to left). Showalter and Hayes-Harb (2015) accounted for this in a second experiment in which they provided participants with training on the direction in which the Arabic script is read and indicated the word-initial graphemes of stimuli with an arrow on several example slides. The experiment was repeated with $N = 8$ participants but again no advantage was found for the Arabic script group on word learning or phonemic contrast acquisition (Showalter & Hayes-Harb, 2015).

Mathieu (2016) explored the effects of an unfamiliar script on the acquisition of an Arabic phonemic contrast in early vocabulary learning and found a negative effect for presenting word forms in Arabic. Native English speakers were shown stimuli in one of four script conditions, Arabic, Cyrillic, a Roman/Cyrillic blend, or a no orthography condition. Stimuli were real and pseudowords in Arabic which began with either a pharyngeal /ħ/ or uvular fricative /χ/. They were paired with line-drawn images and Arabic native-speaker audio recordings. Participants learned six minimal pairs in a learning phase, then took a criterion test featuring only pictorial and auditory stimuli that they repeated until achieving 90% accuracy. After passing the criterion test, they performed a phonological contrast test by responding Yes or No to match and mismatch items.

No significant effect of script condition was found on accuracy, but paired comparisons revealed individuals in the Arabic script condition performed significantly worse than participants in the no orthography condition in the phonological test phase. This was taken as evidence of the hindrance a foreign script can cause in early L2 phonological acquisition, which in turn can impact on the learning of spoken word forms.

Nonetheless, the mean number of cycles required for participants who saw the Arabic script to pass the criterion word-learning test was $M = 2.5$ cycles, compared to $M = 2.8$ cycles for the no orthography condition and $M = 3$ cycles for the Cyrillic group. This suggests that as opposed to a detrimental effect, there may actually have been some benefit for participants who received the Arabic script as they required fewer exposures to accurately match auditory word forms with meaning. Additionally, the researchers had predicted the Cyrillic-trained group would outperform the Arabic-trained group given not all Cyrillic graphemes would be unfamiliar; however, this was not the case.

A similar result was found in a study by Hayes-Harb and Cheng (2016) in which participants who completed a word learning task where written input was provided in Zhuyin, a completely unfamiliar script, outperformed those who received written forms in the more familiar Pinyin, a writing system that uses the Roman alphabet.

It is also important to consider the nature of the written and spoken input provided in Mathieu's (2016) study. Neither of the target phonemes existed in the participants' L1 phonemic inventory. Additionally, in the Arabic script condition, the word-initial phoneme was represented by either the letter ح or خ. These two graphemes share the same basic shape and are distinguishable only by the presence or lack of a dot. Laufer (1988) underscores that words with similar forms can increase difficulty for learners. Thus, the learning task presented a set of low-variability visually and phonologically similar items, as in Showalter and Hayes-Harb (2015).

Adwan-Mansour and Bitan (2017) looked at variability in a stimulus set and its effect on learning for adults studying a language with an unfamiliar writing system. Researchers trained L1 Hebrew participants on a set of six letters from an artificial language and either a set of words with low or high variability. Results revealed the group that received the high-variability words were more accurate in identifying correct transcriptions for both trained and untrained words. Low variability in the stimulus set led to poorer performance in word identification.

Adwan-Mansour and Bitan (2017) may have operationalized variability using the number of items in the stimulus set but there are other ways to go about it. Learning benefits have also been found for high variability in phonemic and graphemic (Apfelbaum et al., 2012), morphological (Tamminem et al., 2015) and acoustic input (Sommers & Barcroft, 2011), for example, talker gender and number. It is also interesting to note Adwan-Mansour and Bitan (2017) conducted a letter-training phase (i.e., explicit instruction in grapheme-phoneme correspondences) prior to presenting participants with a set of stimuli to learn.

**Letter Training Studies**

Studies comparing participants who receive explicit letter training to those who are trained implicitly using whole-words have found that letter training results in greater accuracy in word recognition and generalization of learning to novel word forms (Bishop, 1964; Bitan & Booth, 2012; Bolger, 2007; Brennan & Booth, 2015).

In an early study of the effects of training whole words versus individual letters on decoding skills in reading, Bishop (1964) trained native English speakers on either Arabic letters or Arabic pseudowords. In the letter training, graphemes were printed on cards and presented for four seconds while an audio recording of the corresponding phoneme was played. The performance of both groups on a decoding task for untrained words and a letter-matching task was compared to that of a control group who received no pre-test training. Results showed that the letter-training group significantly outperformed the whole-word and control groups on word reading and correct grapheme-phoneme identification. The letter-training group identified $M = 11.85$ of 12 letters, compared to $M = 8.05$ by the whole-word group and $M = 6.75$ by the control group.

While Bishop's (1964) study is relevant to the current research, as it provides an example of a successful letter training in Arabic delivered to L1 English absolute beginners, the task participants performed was focused on testing the skill of decoding, not on word learning. In fact, the main aim of the study was to provide insight into approaches to whole-word or phonics-based reading instruction in English. Word cards featured written forms with no images, where Arabic graphemes were presented left to right, following English reading direction. As a result, participants did not learn the meaning of the stimulus items, only how to associate their written and spoken forms.

In Bolger (2007), L1 English native speakers received either letter training for unfamiliar graphemes, or whole-word training for English pseudowords written in in the Korean Hangul script. The letter-training group outperformed the whole-word group in a productive measure of decoding and a receptive measure of word learning. They also had significantly greater accuracy on a series of delayed post-tests, showing enhanced retention of learned word forms. Bolger posited that this was due to their increased attention and focus on individual word components during the word learning task. Research has shown that the amount of processing can affect the degree to which an L2 learner is cognitively involved in a learning task, and enhance long-term memory for new word forms (Laufer & Hulstijn, 2001; Schmitt, 2008).

In their studies Bishop (1964) and Bolger (2007) used authentic orthographic systems, while Bitan and Karni (2003) created a Morse-like code of symbols to test if whole-word training could lead to implicit learning of grapheme-phoneme correspondence. Participants were trained on sets of pseudowords under three conditions, 1) whole-word training, 2) letter training followed by whole-word training, and 3) training in an arbitrary condition with non-alphabetic pictographs. Results showed that only the letter training led to declarative grapheme-to-phoneme knowledge. Moreover, unlike in the whole-word or arbitrary conditions, under the letter-training condition participants scored at above chance on the first block of a translation task.

Brennen and Booth (2015) also looked at instruction in an artificial language both at the letter and word unit on processing speed and sensitivity to graphemes. They trained participants in either small grain (i.e., individual letters) or large grain (i.e., letter strings) units and a set of words. Participants who received the letter training had significantly longer response times across all measures. Researchers believed this may have been due to delays caused by bottom-up processing (i.e., decoding of individual graphemes), versus a more top-down (i.e., whole-word focus) approach to word identification. This explanation is in line with the view that bottom-up processing can lead to processing delays in L2 reading (McLeod & McLaughlin, 1986).

In summary, previous research has found no benefit (Showalter & Hayes-Harb, 2015) or a detrimental effect (Mathieu, 2016) when Arabic written forms are provided to absolute beginners in tasks related to phonemic contrast perception and word learning. However, participants in these studies did not receive any initial letter training in Arabic. At the same time, studies with English native speakers have shown that letter training can lead to enhanced performance in decoding and letter identification in Arabic (Bishop, 1964), and in other authentic and artificial languages (Bitan & Booth, 2012; Bitan & Karni, 2003), as well as facilitate retention in a learning task (Bolger, 2007).

**Present Study and Research Aims**

The aim of the current study was to explore the potential benefit of letter training on subsequent word learning. Letters for training were selected to be visually different, were always presented in the same form (i.e. there was no allographic variation) and were unconnected to ensure that letters could be identified as separable units. Performance on a word learning task in terms of speed and accuracy was compared for two groups: one that received letter training, and control group that did not. It was predicted learning Arabic grapheme-phoneme correspondences would assist the letter-training group in learning the target words. As in Bitan and Karni (2003), it was also expected that an early benefit for participants who received the letter training would emerge in the initial block of word presentation. Furthermore, slower response times were expected for the letter-training group, given the enhanced processing burden of decoding (Brennan & Booth, (2015).

Two delayed measures of learning were conducted five to seven days after the original session. Following Bolger (2007), it was predicted that the letter-training group would show enhanced retention for learned input, and outperform the control group on the delayed measures of learning. An additional line of investigation centered around how many letters in the input needed to be familiar (i.e., trained) for a letter training group to enjoy a word-learning benefit.

The research questions for the current study are:

1. Is there a benefit to providing letter training in an unfamiliar script (Arabic) for *immediate* word recognition and passive recall tasks?
2. Is there a benefit to providing letter training in an unfamiliar script (Arabic) for *delayed* word recognition and passive recall tasks?
3. Is letter training more beneficial for words containing more taught letters?

## Method

### *Participants*

Forty-four adult native English speakers participated in the study. Participants were all students at the University of Nottingham, UK who had no previous experience learning Arabic, Hebrew or a language written in the Arabic script. In order to take part in the study, it was a requirement that individuals had not spent any length of time in an Arabic speaking country or had significant exposure to written or spoken Arabic. Participants were randomly assigned to a Letter Training (LT) or No Letter Training (No-LT) group and were paid £15 pounds for taking part in the study.[1]

All participants completed a language background questionnaire that asked about their gender, age, level of education, experience with Arabic, and previous experience with foreign language study, including languages learned and an estimate of their highest achieved proficiency in each language (see Table 5.1). This data was collected to help investigate the impact of previous experience achieving higher levels of fluency in a second or foreign language, learning non-European languages, and studying different script languages on the Arabic letter training and word learning tasks.

[1] Funding for the study was provided by a research grant from the Qatar Foundation International.

There were $N = 17$ male and $N = 25$ female participants. Participants ranged in age between 18 and 49 years. All participants held or were working toward a degree at university ($N = 25$ undergraduate, $N = 13$ at master's, and $N = 4$ at postgraduate). All participants reported having studied a foreign language to a beginner or intermediate level. The most reported languages were French ($N = 26$), Spanish, ($N = 17$), German, ($N = 10$), and Mandarin ($N = 7$), with a number of participants having experience studying a language with a different script ($N = 17$).

**Table 5.1**
*Participant Demographic and Language Background Information*

| Measure | Letter training group | | No letter training group | | Full sample | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Age | 22.67 | 6.61 | 24.57 | 8.05 | 23.62 | 7.34 |
| Number of languages studied | 2.33 | 1.35 | 1.95 | 0.67 | 2.14 | 1.07 |
| Highest proficiency level reached | 1.81 | .81 | 1.90 | .83 | 1.86 | 0.81 |

*Note*. Participants were asked to rate their proficiency on a 3-point scale (1 = beginner, 2 = intermediate, 3 = advanced) in a language other than English.

Data from two participants were excluded: one because the participant achieved less than 50% accuracy on the letter training test; and the other because it was discovered that the participant had significant prior experience with the Arabic script. This left the LT and No-LT groups with 21 participants each. One participant from the LT group was not able to return for the second session. Thus, data for delayed measures reflects 20 LT and 21 No-LT participants. This participant's data has not been included in Table 5.1.

## Materials

### *Letters*

Eight Arabic consonants were chosen for the letter training, see Table 5.2. Arabic is a cursive script where letters take different forms depending on their place at the beginning, middle, or end of a word, or their status as connecting or non-connecting letters. Only the isolated letter shapes were used. This was done to ensure there was no allographic variation and the individual letters would be easy to identify. The letters were referred to by their spoken letter names (e.g. ل as 'lam') given that this is how letters are most commonly taught in Arabic as a Foreign Language (AFL) materials that introduce the alphabet to beginners. Arabic is also characterized by acrophony in which a letter name begins with the sound the letter makes (Tibi et al., 2020). All the taught phonemes exist in English and the isolated letter forms were distinct except for the ر */r/* and ز */z/* which share the same base letter shape and are only distinguished by the presence or lack of a dot above the letter. Because of their similarity, these two letters were never paired in any of the two-forced-choice tasks (e.g. choose ر or ز).

**Table 5.2**

*Written Stimulus Material from the Letter Training Session and Arabic Word Learning Task*

| Letters | | | | | | | |
|---|---|---|---|---|---|---|---|
| ب | ر | ز | ف | ك | ل | م | ن |
| /b/ | /r/ | /z/ | /f/ | /k/ | /l/ | /m/ | /n/ |
| ba | ra | zay | fa | kaf | lam | meem | noon |

| Words |
|---|

| Type 1 | Type 2 | Type 3 | Type 4 |
|---|---|---|---|
| all-FAM | un-FAM1 | un-FAM2 | no-FAM |
| ف ل ف ل | ن ه ر | م ق ص | خ س |
| /filfil/ | /nahr/ | /miqasˤ/ | /xas/ |
| *pepper* | *river* | *scissors* | *lettuce* |
| ف م | ف ط ر | م ش ط | ع ش |
| /fam/ | /fitˤr/ | /miʃtˤ/ | /ʕuʃ/ |
| *mouth* | *mushroom* | *comb* | *nest* |
| م ف ك | ج ز ر | ف ن د ق | ق و س |
| /mifak/ | /dʒazar/ | /funduq/ | /qaws/ |
| *screwdriver* | *carrots* | *hotel* | *bow* |
| م ر ك ب | ك ر ك د ن | ف ه د | ط و ق |
| /markib/ | /karkadan/ | /fahd/ | /tˤawq/ |
| *boat* | *rhinoceros* | *leopard* | *necklace* |
| ك ر ز | ج ب ل | س ط ل | س د |
| /karaz/ | /dʒabal/ | /satˤl/ | /sad/ |
| *cherry* | *mountain* | *pail* | *lock* |
| ك ل ب | ق م ر | ج س ر | س ي ا ج |
| /kalb/ | /qamar/ | /dʒisr/ | /sijaːdʒ/ |
| *dog* | *moon* | *bridge* | *fence* |
| ز ر | س م ك | د ف ت ر | ش ا ط ئ |
| /zir/ | /samak/ | /daftar/ | /ʃaːtˤaːʔ/ |
| *button* | *fish* | *notebook* | *beach* |

*Note*. Arabic words were presented to participants with a forced spacing manipulation, to facilitate letter-form recognition, and with images instead of written English definitions to represent meaning.

### *Words*

Twenty-eight concrete Arabic nouns were selected as stimuli for the word learning task (see Table 7.2). Stimuli were divided into four types. Type 1 (all-FAM) consisted of words only containing familiar letters (i.e., letters which were included in the

Letter Training), Type 2 (un-FAM1) contained words with one unfamiliar letter, and Type 3 (un-FAM2) contained words with two unfamiliar letters. Type 4 (no-FAM) was made up of seven control items in which none of the component letters were familiar. For words with two unfamiliar letters, a familiar letter always appeared in the word-initial or word-final position.

Arabic has three long vowels and three short vowels. Long vowels are represented by the letters أ /*ā*/, و /*ū*/, and ي /*ī*/ and short vowels are either unwritten or denoted with diacritic marks. To avoid confusion about the presence or absence of vowels, no long vowels were taught and all-FAM, un-FAM1 and un-FAM2 words featured only unwritten short vowels. The word-types were balanced for syllables and length. All words were either 1 or 2 syllables long, with the exception of the 3-syllable word كركدن *karkdan* that was used to satisfy the familiar letter constraints of Type 2 items. All words contained between 2 and 5 graphemes ($M = 3.07$, $SD = .77$).

Similar to Bishop (1964), a script manipulation was used to ensure recognition of the taught graphemes and facilitate decoding (Dai et al., 2013) and recognition of the target Arabic words. A space was inserted between all graphemes, such that no words contained connected forms. For example, the word فلفل *filfil* was written as ف ل ف ل.

*Images & Audio*

Target words, with native-speaker audio recordings, were paired with color images from the book *First Thousand Words in Arabic* (Amery, 2014). This textbook is an Usborne self-study course aimed at beginner English speaker learners and thus covers vocabulary at a novice level. Audio stimuli were WAV files and consisted of the word read aloud once by a female talker. Pictorial stimuli were color line drawings on a white background. Researcher-added arrows were used to clarify the meaning for three images.

Images were normed for their name agreement (i.e., the image inducing the intended label), age of acquisition, frequency of use/encounter, imageability, and visual complexity/image detail, by a separate set of participants. Participants in the image norming study ($N = 18$) were native English speaker students enrolled at the University of Nottingham who received course credit for their participation. They were shown items using a computer-based online survey hosted by Bristol Online Survey (BOS) and asked to identify the image, then respond to a series of open-ended and Likert scale questions.

The information statistic $H$ was computed for each image as a measure of name agreement that takes into account both the percentage of agreement and distribution of names given (Snodgrass & Vanderwart, 1980). There was high name agreement among participants with 16 of the 28 images eliciting an $H$ value of .0 (i.e., perfect name agreement) and an overall mean of $M = .53$ $SD = .76$. Age of acquisition results showed most words were known to participants by or before the age of five ($M = 4.55$ $SD = 1.23$). Frequency of use and encounter indicated that the words were generally familiar (on a 5-point scale, use $M = 3.77$ $SD = 1.18$, encounter $M = 4.05$ $SD = 1.16$). Imageability was high across all items (on a 7-point scale, $M = 6.19$ $SD = .50$), indicating it was considered "easy" to generate a mental image for the words. Ratings of visual complexity suggest that the images contained a similar level of detail (on a 7-point scale, $M = 3.79$ $SD = .78$). There were several outliers of interest, notably the image for a *lock (water navigation)* was on average acquired by participants at a later age and both *lock* and *hotel* elicited a greater range of names (e.g., restaurant and cinema for the image of a hotel).

Nonetheless, overall results suggested that the images used to represent word meaning in the main study would be generally familiar to participants and that they had similar properties on features that could influence response times.

*Letter Training*

The letter training was run using E-Prime in the University of Nottingham Psycholinguistics and Language Learning Lab. It took roughly five minutes to complete. The training was done in three rounds in which each of the 8 graphemes was presented in isolation (see top of Table 5.2). In the first round, participants saw a grapheme while an audio file repeated the letter name three times. In the second round, participants saw the grapheme and heard two repetitions of a word that began with the corresponding phoneme. In the third round, they saw the grapheme and heard two repetitions of both the letter name and example word. Audio files for the letter training phase were recorded by a male Arabic native speaker talker. This was a different talker to the one used to make the recordings for the word learning task. Example words used in the letter training were distinct from the words presented in the word learning task and were *not* written on the screen. Participants could move through the letter training at their own pace (i.e. they pressed a button on a keyboard to advance to the next letter), but the audio files could not be replayed.

At the end of the three rounds, participants took a test in which two graphemes appeared on screen and an audio recording of a letter name was played. They were asked to click on the letter they heard. After the click, the correct grapheme became green regardless of whether they had replied correctly or incorrectly. There were 16 test items.

Each grapheme appeared correctly in two test items with different distractors. Note that ر */r/* and ز */z/* were never paired due to their visual similarity. A percentage correct score was shown at the end and only participants who scored 75% or higher were permitted to continue to the word learning task.

*Word Learning*

Participants learned the target words on the same computer using E-Prime. All participants were shown a set of on-screen instructions which explained that Arabic is read right to left. An example word was displayed in which the rightmost letter was marked with an arrow and shown in green while the word was read aloud. This was also done by Showalter and Hayes-Harb (2015) to ensure participants were able to map the word-initial phoneme to the correct Arabic grapheme. Participants were told they would see a picture on the screen while seeing and hearing a word. The written words were presented in the Arabic script with letter spacing. They were asked to decide whether the image depicted the word that was shown and heard. They were given on-screen feedback "Correct" or "Incorrect" based on their response. Participants indicated their responses with a key for "Yes- the picture matches the written and audio word" or a key for "No – the picture doesn't match the written and audio word." As the participants had no prior knowledge of Arabic, they were instructed to use the on-screen feedback to learn the words.

Learning took place over three blocks. In each block, a word was presented once in a match and once in a mismatch condition. Participants were shown their percent correct responses at the end of each block. Across the blocks there were 28 match pairs, 84 mismatch pairs and 168 randomized trials. The mismatch pairs were created by pairing a written form and audio-word with a picture of another word from the same type (e.g., all-FAM Arabic written and audio word for *pepper* paired with mismatch all-FAM image of *boat*). Participants were never exposed to a mismatch pair more than once.

*Word Recognition Task*

The third block of the word learning was used as a word recognition task/measure (i.e., participants recognized, or not, the correct meaning/image associated with an audio and written word form).

*Recall Task*

A recall task was carried out as a measure of the strength of participants' word knowledge, as passive recall indicates a stronger memory trace than passive recognition alone (Laufer & Goldstein, 2004). Participants heard the audio file of a target word and were simultaneously shown the word's written form presented using the Arabic spaced script in E-Prime. They used the keyboard to either type the English meaning of the word or write "skip" if they did not know it. They were told they could also describe the associated picture and that they would not be penalized for English spelling mistakes in their responses. In contrast to a word recognition task in which participants indicate yes/no if a word corresponds to a meaning/image, a recall task requires participants to demonstrate their understanding of a word through production of an L1 translation equivalent (i.e., participants recognized, or not, the written and audio form of a word and were able to recall its meaning in English). The task consisted of 28 trials with each target word appearing once in random order.

*Delayed Word Recognition and Recall*

A delayed word recognition task and a recall task were carried out five to seven days after the initial session. The delayed word recognition task was a fourth block of the word learning, containing 56 randomized trials with 28 match and 28 novel mismatch pairs. The delayed recall task was the same as the immediate one and consisted of 28 randomized trials.

*Procedure*

The experiment had two sessions. In Session 1, participants in the Letter Training group started by learning a set of Arabic letters before proceeding to the word learning task. The No Letter Training group began with the word learning task. Both groups of participants saw the same introductory and instruction screens before beginning the learning task. The learning task consisted of three blocks. Participants could take breaks between blocks if needed (although most chose not to). The third block of the learning task was used as a measure of word recognition. Following the third block, participants were given a brief language background questionnaire. They then finished the session with the recall task.

Session 2 took place five to seven days after Session 1. It began with a fourth block of the learning task, which served as a delayed measure of word recognition. This was followed by a delayed recall task. Participants had no exposure to the words in the days between the sessions. Session 1 took between 30 and 35 minutes for participants who completed the letter training and between 25 and 30 minutes for those who did not take part in the letter training. Session 2 was approximately 10-15 minutes long.

**Results**

The focus of the study was investigating the effect of letter training on word learning. To do this, performance by the LT and No-LT groups must be compared. Not all of the participants who received the letter training achieved 100% accuracy in mapping the eight taught letters to their corresponding sounds (i.e. mastered the grapheme-to-phoneme correspondences). The distinction between those who achieved 100% accuracy, and those that did not, may be important. Therefore, it was decided to treat the groups separately in the analyses. These two groups will be referred to as the LT-High and LT-Low groups respectively.

There are a number of results to report. The results of the letter training (LT), which half of the participants took part in, are presented first. For this only the LT groups are relevant (LT-High and LT-Low groups). Following this are the analyses of word recognition and recall for both the immediate and delayed tasks, as well as for the word type data. For these analyses, there are three participant groups: LT-High, LT-Low, and No-LT.

All analyses were conducted using SPSS 24.0.0.2.

*Letter Training*

Mean accuracy and response times (in milliseconds) for the letter training are reported by participant group and by letter in Table 5.3. Of the 21 participants who took part in the training, eleven achieved 100% accuracy. The remaining ten had scores ranging from 75% - 94%, with 75% being the minimum score required for continuing to the word-learning phase. Independent samples *t*-tests confirmed the LT-High group was more accurate, $t(19) = 9.464$, $p < .001$, than the LT-Low group, but the two groups did not differ in their response times, $t(19) = .122$, $p = .904$.

**Table 5.3**

*Means and Standard Deviations of Accuracy and Response Times for the Letter Training*

| By Group | | | | | |
|---|---|---|---|---|---|
| Group | | Accuracy | | Response time | |
| | *N* | *M* | *SD* | *M* | *SD* |
| LT-High | 11 | 1.0 | .00 | 2846.98 | 287.47 |
| LT-Low | 10 | .85 | .05 | 2811.07 | 927.06 |
| Both | 21 | .93 | .08 | 2829.88 | 654.53 |

| By Letter | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Letter | | Accuracy | | | | Response time | | | |
| | | LT-High | | LT-Low | | LT-High | | LT-Low | |
| | | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| ب | /b/ | 1.00 | .00 | .80 | .26 | 2678.14 | 618.94 | 2449.05 | 890.07 |
| ر | /r/ | 1.00 | .00 | .90 | .21 | 2663.68 | 604.40 | 2957.20 | 2190.74 |
| ز | /z/ | 1.00 | .00 | .70 | .34 | 2972.64 | 1314.20 | 3460.39 | 1011.69 |
| ف | /f/ | 1.00 | .00 | 1.00 | .00 | 3150.64 | 891.78 | 2661.95 | 1640.51 |
| ك | /k/ | 1.00 | .00 | .80 | .26 | 2531.05 | 686.35 | 2608.55 | 1153.71 |
| ل | /l/ | 1.00 | .00 | .80 | .26 | 2752.14 | 586.70 | 2855.55 | 1206.61 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| م | /m/ | 1.00 | .00 | .90 | .21 | 3049.36 | 1159.92 | 2498.75 | 304.23 |
| ن | /n/ | 1.00 | .00 | .90 | .21 | 2978.18 | 993.67 | 2540.60 | 724.36 |

A one-way analysis of variance (ANOVA) was conducted on the LT-Low group to probe the main effect of letter (ن, م, ل, ك, ف, ز, ر, ب) on accuracy. While no main effect was found, $F(7,72) = 1.505$, $p = .179$, inspection of the data in Table 5.3 indicates that some graphemes are easier or harder for participants to learn. For example, ر /r/ and ز /z/ might be more difficult to learn because of their visual similarity. To more fully explore whether there were differences between the graphemes, planned comparisons between all letters were carried out, but only those that reached significance at $p < .05$ are reported. This analysis revealed that participants were significantly less accurate in responding to ز /z/ compared to ف /f/ $p < .05$. Participants were also generally more accurate in their responses to the letter ف /f/ than to all of the other graphemes included in the training, $p < .001$. Notably ف /f/ shares no form components with the other letters. The data was also examined via a series of paired *t*-tests, with comparisons reaching significance at $p < .05$ reported here. The letter ف /f/ was found to be significantly easier to learn than ب /b/, ك /k/, ل /l/ and ز /z/ (*p*'s < .05). No analyses were carried out on the LT-High group who performed at ceiling.

In looking at response time data for the LT-Low participants, a one-way ANOVA showed there was no main effect for letter on response time, $F(7,71) = .656$, $p = .708$. Again, the data was examined via a series of paired *t*-tests, with comparisons reaching significance at $p < .05$ reported here. The letter ز /z/ was responded to significantly slower than six of the other taught letters: ب /b/ $p < .01$, ر /r/ $p < .05$, ف /f/ $p < .05$, ك

/k/ $p < .05$, م /m/ $p < .05$, and ن /n/ $p < .05$. There were no significant differences in the LT-High group's response times by letter ($p > .05$).

*Immediate Word Recognition – Accuracy*

Participants were asked to learn words across three blocks. Only data from match pairs for all-FAM, un-FAM1 and un-FAM2 words was considered (i.e., words containing graphemes introduced in the letter training). Mean accuracy for word recognition measures is reported in Table 5.4.

**Table 5.4**

*Mean Accuracy and Response Times in the Immediate and Delayed Word Recognition Tasks by Participant Group*

Accuracy

| Group | Block 1 | | Block 2 | | Block 3 | | Block 4/Delayed | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| LT-High | .76 | .14 | .78 | .11 | .92 | .10 | .89 | .10 |
| LT-Low | .70 | .11 | .74 | .13 | .82 | .10 | .77 | .11 |
| No-LT | .63 | .12 | .78 | .14 | .82 | .08 | .80 | .12 |

Response times

| Group | Block 1 | | Block 2 | | Block 3 | | Block 4/Delayed | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| LT-High | 2571.18 | 783.45 | 1872.50 | 556.87 | 1605.40 | 479.93 | 1596.87 | 396.56 |
| LT-Low | 2404.05 | 595.72 | 1726.62 | 451.66 | 1621.36 | 307.38 | 1477.57 | 216.23 |
| No-LT | 2424.95 | 690.47 | 1960.08 | 622.71 | 1546.69 | 364.13 | 1620.63 | 315.87 |

*Note.* Block 3 was used as the immediate word recognition measure. Block 4 indicates the delayed word recognition task.

**Table 5.5**

*One and Two-Way ANOVA Statistics for Immediate and Delayed Measures*

Immediate

| Measure | ANOVA | | | | | | |
|---|---|---|---|---|---|---|---|
| | | By subject | | | By item | | |
| | Effect | $F_1$ ratio | *df* | η2 | $F_2$ ratio | *df* | η2 |
| Word recognition - accuracy | G × B | 3.023* | 3.485, 67.958 | .134 | 3.112* | 4,120 | .094 |
| Word recognition - response time | G | .357 | 2,39 | .018 | 2.926 | 2,60 | .089 |
| | B | 48.552*** | 1.644,64.122 | .555 | 98.744*** | 1.687,101.221 | .622 |
| | G × B | .631 | 3.288,64.122 | .031 | 1.080 | 3.374,101.221 | .035 |
| Recall - accuracy | G | 9.339*** | 2,39 | .324 | 9.715*** | 2,60 | .245 |

Delayed

| Measure | ANOVA | | | | | | |
|---|---|---|---|---|---|---|---|
| | | By subject | | | By item | | |
| | Effect | $F_1$ ratio | *df* | η2 | $F_2$ ratio | *df* | η2 |
| Word recognition - accuracy | G | 4.494* | 1,38 | .191 | 3.024 | 2,60 | .092 |
| | T | 6.680* | 1,38 | .150 | .225 | 1,60 | .004 |
| | G × T | .507 | 2,38 | .026 | 2.085 | 2,60 | .065 |
| Word recognition - response time | G | .085 | 2,38 | .004 | .554 | 2,60 | .018 |
| | T | .150 | 1,38 | .700 | 3.055 | 1,60 | .048 |
| | G × T | .976 | 2,38 | .049 | 1.107 | 2,60 | .036 |
| Recall - accuracy | G | 11.298*** | 2,38 | .373 | 10.433*** | 2,60 | .258 |
| | T | 7.056** | 1,38 | .157 | 12.062** | 1,60 | .167 |
| | G × T | .411 | 2,38 | .021 | .339 | 2,60 | .011 |

*Note.* ANOVA = analysis of variance; G = group; B = Block; T = Time-point.
*$p < .05$. **$p < .01$. ***$p < .001$.

A 3x3 two-way ANOVA evaluated accuracy with group (LT-High, LT-Low, No-LT) as a between-subject factor and block (Block 1, Block 2, Block 3) as a within-subject factor. Data were normally distributed and homogeneity of variances ($p > .05$) and covariances ($p > .05$) was found using Levene's test of homogeneity of variances and Box's tests respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction by subject, $\chi 2(2) = 6.08$, $p = .048$, and consequently the Greenhouse-Geisser score is reported. As can be seen in Table 5.5, there was a significant interaction between letter-training group and block both by subject and by item. Pairwise comparisons revealed the LT-High group was more accurate than the No-LT group in Block 1 ($p < .05$). Critically, the LT-High group outperformed both the No-LT ($p < .05$) and LT-Low ($p < .05$) groups in Block 3, which was the immediate measure of word recognition.

Mean accuracy varied for each group across block. Planned comparisons revealed participants from the LT-High group were more accurate in Block 3 ($p < .01$) after having received a second exposure to the words in Block 2, but made no significant accuracy gains between Blocks 1 and 2 ($p >. 05$). This was different from the No-LT group who made early gains with greater accuracy in Block 2 ($p <. 001$) than in Block 1, but had no difference in mean accuracy in the final two blocks ($p >. 05$). Words learned by the LT-Low group were responded to more accurately in Block 3 than in Block 1 ($p <. 05$), but this was significant by items only. No other pairwise comparisons were significant ($p$'s $> .05$).

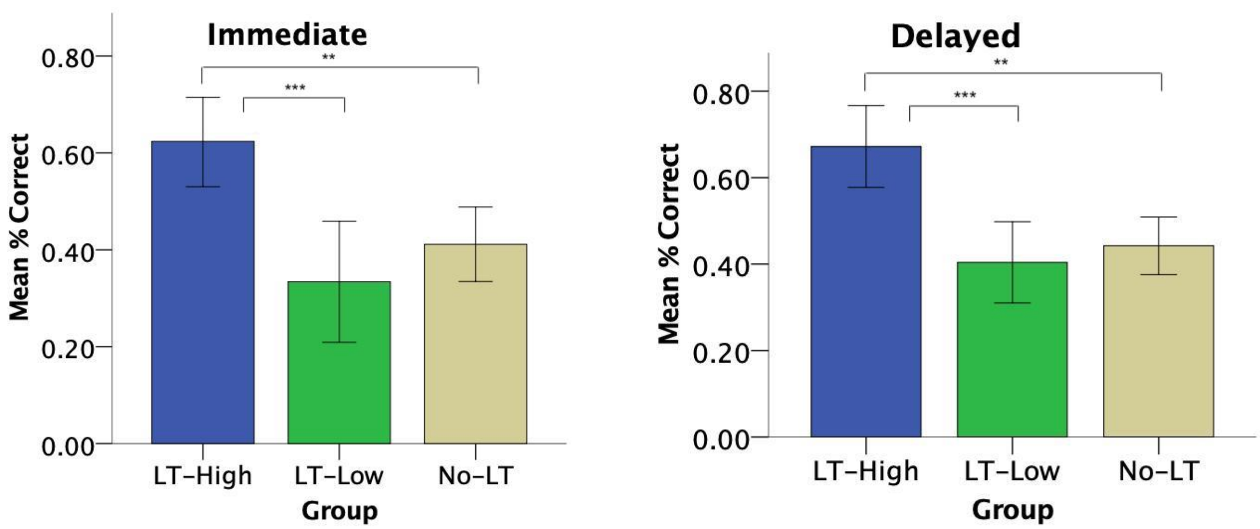
*Immediate Word Recognition – Response times*

Response times that were 2.5 standard deviations above or below the mean in the recognition task were removed (less than .01% of the data). Mean response times for word recognition measures are reported in Table 5.4. Response times were *only* considered in the word recognition task, as potential differences in typing skill would likely influence response times for the recall task.

A 3x3 two-way ANOVA evaluated response time with group (LT-High, LT-Low, No-LT) as a between-subject factor and block (Block 1, Block 2, Block 3) as a within-subject factor. Homogeneity of covariances was not found ($p > .05$), as assessed by Box's test of equality of covariance matrices. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction by subject, $\chi2(2) = 9.27$, $p = .010$, and by item, $\chi2(2) = 12.107$, $p = .002$, thus Greenhouse-Geisser scores were reported. As expected, there was a main effect of block by subject and by item, see Table 5.5. Participants became faster with more exposures to the words. Pairwise comparisons revealed they were significantly faster to respond in Block 2 than in Block 1 ($p < .001$), and in Block 3 than in Blocks 2 ($p < .01$) and 1 ($p < .001$). No main effect was found for group by subject or by item and there was no statistically significant two-way interaction.

**Figure 5.1**

*Mean Accuracy in the Immediate (Block 3) and Delayed (Block 4) Word Recall Tasks by Participant Group*




*Note.* $^{**}p < .01$. $^{***}p < .001$. Error bars show standard errors: Confidence interval 95%.

*Immediate Recall – Accuracy*

A one-way ANOVA was performed to determine the effect of letter-training on accuracy in the immediate recall task, see Figure 5.1. There were no outliers in the data, as assessed by inspection of a boxplot. Accuracy was normally distributed for the three groups, as assessed by Shapiro-Wilk's test (p > .05), and there was homogeneity of variances, as assessed by Levene's test for equality of variances (p > .05). As reported in Table 5.5, results showed a significant main effect of group by subject and by item, with large effect sizes. Planned comparisons indicated the LT-High group outperformed the No-LT group ($p < .01$), and the LT-Low group ($p < .001$) on the immediate recall task.

*Delayed Word Recognition – Accuracy*

Participants took part in a fourth block of the word learning task in Session 2, which occurred five to seven days after Session 1. Mean accuracy for the delayed measures are reported in Table 5.4. A 3x2 two-way ANOVA evaluated accuracy with group (LT-High, LT-Low, No-LT) as a between-subject factor and time-point (Immediate- Block 3, Delayed- Block 4) as a within-subject factor. There were no outliers, as assessed by a boxplot and data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). Homogeneity of variances was present ($p > .05$), as assessed by Levene's test of homogeneity of variance. Homogeneity of covariances as assessed by Box's test of equality of covariance matrices was found by subject (p > .05), but not by item ($p = .001$). The ANOVA results are reported in Table 5.5. There was a main effect of group; the LT-High participants outperformed the No-LT ($p < .05$) and LT-Low participants ($p < .05$). However, this was significant by subjects only, and failed to reach significance by items. There was also a main effect for time-point which reached significance by subjects but not by items. Participants in the LT-Low group were significantly less accurate on the delayed measure than the immediate one ($p < .05$). No other pairwise comparisons were significant ($p$'s > .05). There were no significant interactions between group and time-point by subject or by item.

*Delayed Word Recognition – Response Times*

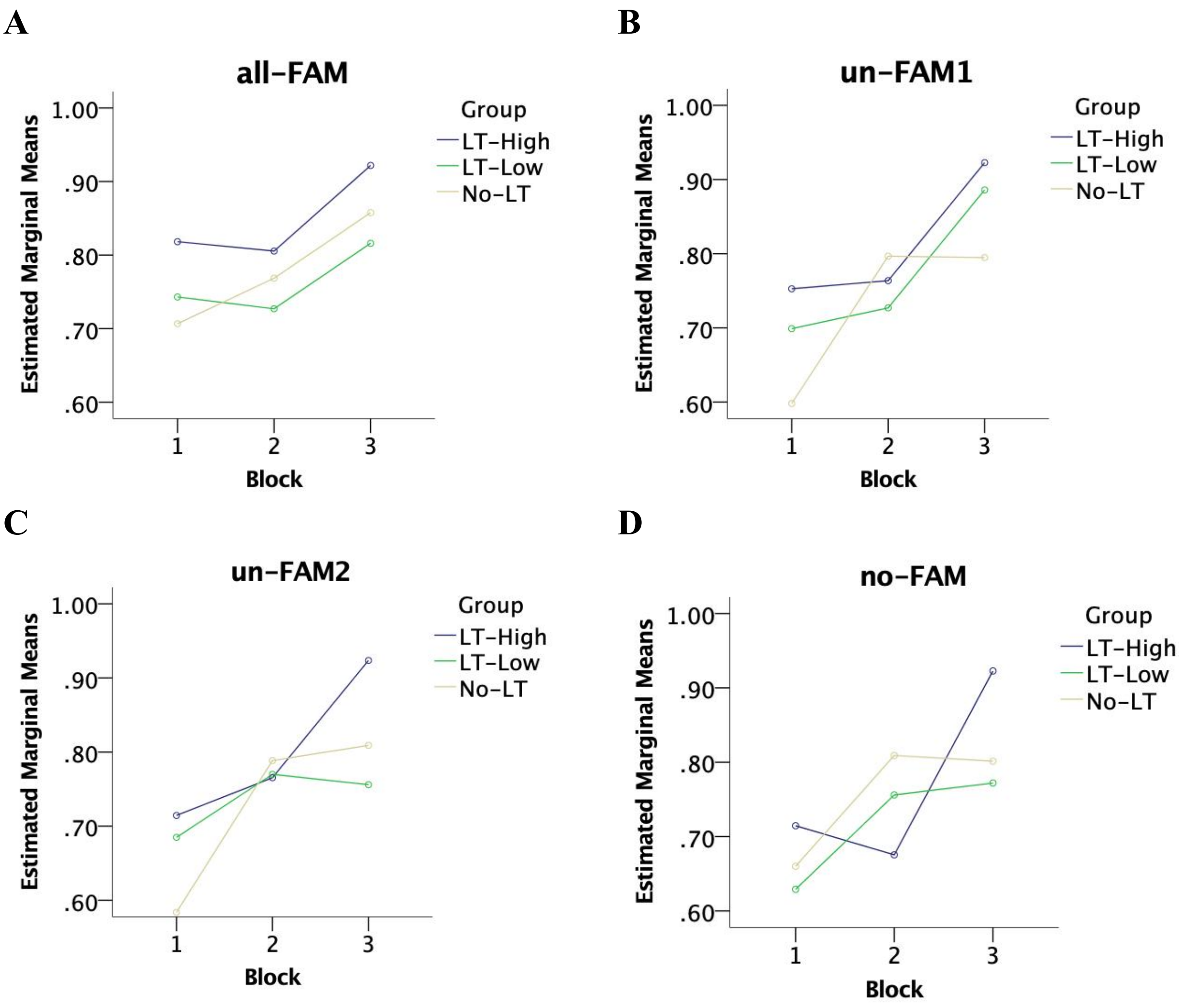
A 3x2 two-way ANOVA evaluated mean response times with group (LT-High, LT-Low, No-LT) as a between-subject factor and time-point (Immediate-Block 3, Delayed-Block 4) as a within-subject factor. There were no outliers, as assessed by a boxplot and the data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). As can be seen in Table 5.5, no significant interaction was found between letter-training group and time-point by subject or item. There was no main effect of group; participants from different letter-training groups were not significantly faster or slower in their responses. There was also no main effect of time-point; participants were not significantly faster or slower on the immediate or delayed measures.

*Delayed Recall – Accuracy*

The delayed recall task was identical to the immediate task but was given in Session 2. A 3x2 two-way ANOVA evaluated accuracy with group (LT-High, LT-Low, No-LT) as a between-subject factor and time-point (Immediate, Delayed) as a within-subject factor. There were no outliers, as assessed by a boxplot and the data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity of variances, as assessed by Levene's test ($p > .05$), as well as homogeneity of covariances, as assessed by Box's test of equality of covariance matrices ($p > .05$). ANOVA results are reported in Table 5.5. A main effect was found for group by subject and by item. As can be seen in Figure 5.1, the LT-High group outperformed the No-LT group ($p < .01$) and the LT-Low group ($p < .001$). A main effect was also found for time-point by subject and item. Participants were significantly more accurate ($p < .05$) on the delayed than the immediate measure of word recall. No significant interaction effect was found.

**Figure 5.2**

*Mean Accuracy in Word Learning for Word Types: all-FAM, un-FAM1, un-FAM2, no-FAM*




*Note.* Panel A: all-FAM = all familiar letters; Panel B: un-FAM1 = one unfamiliar letter; Panel C: un-FAM2 = two unfamiliar letters; Panel D: no-FAM = no familiar letters (control words).

*Word Type (i.e., Number of Taught Letters) - Immediate Word Recognition Accuracy*

Word type data includes performance on all four categories of words, including control item match-pairs, (i.e., words that contained no taught graphemes). Data was inspected carefully before analysis. Figure 5.2 shows participant groups who took part in the letter training patterned differently for words made up of letters which were mostly familiar (i.e., included in the training), compared to control words and items with two unfamiliar letters. It was thus decided to proceed with the analysis using a mostly Familiar Letters (FAM) category (i.e., all-FAM all familiar letters and un-FAM1 one unfamiliar letter words; $N = 14$), and a mostly Unfamiliar Letters (unFAM) category (i.e., un-FAM2 two unfamiliar letters and no-FAM no familiar letter control words; $N = 14$).

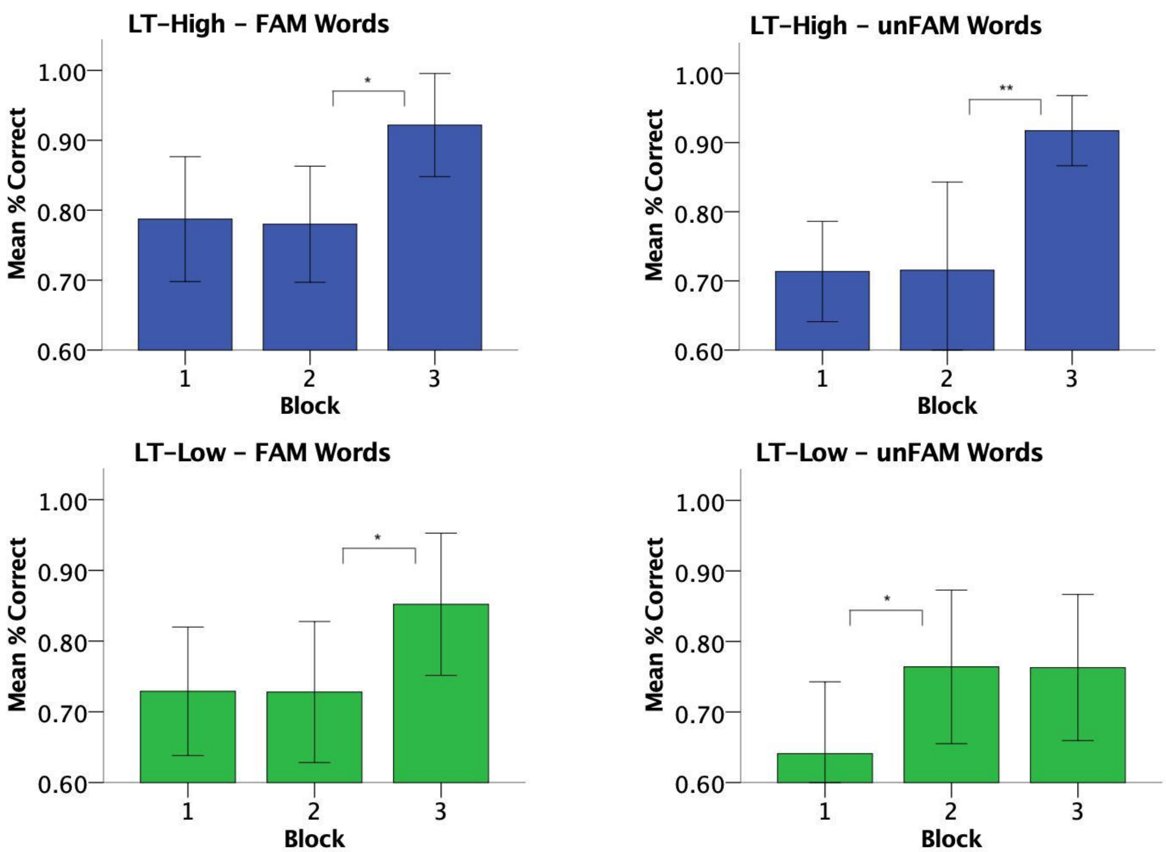
Word type analyses for all measures are reported in Table 5.6. A 3x2x3 three-way ANOVA evaluated accuracy with group (LT-High, LT-Low, No-LT) as a between-subject factor and word type (FAM, unFAM) and block (Block 1, Block 2, Block 3) as within-subject factors. The assumption of homogeneity of variances was met, as assessed by Levene's test for equality of variances ($p > .05$). Mauchly's test of sphericity indicated that the assumption of sphericity was met for the block and word-type interaction, $\chi2(2) = 1.024$, p = .599. A main effect of word type was found, but this effect was only approaching significance by-items ($p = .051$). Pairwise comparisons indicated participants were more accurate on words with more familiar letters ($p < .05$).

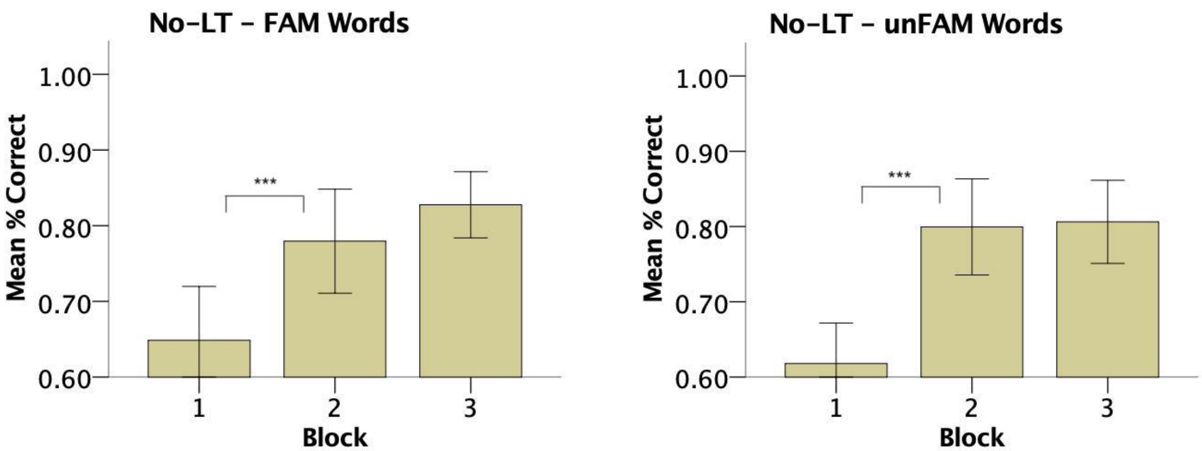
A significant two-way interaction was also found between block and group by subject ($p = .005$), and by item ($p = .002$). Groups performed differently in the initial learning block on words that were made up of mostly taught letters ($p =.039$). The high-performing letter-training group were more accurate than the group who received no training ($p < .05$). No difference in the

Block 1 performance between groups was observed for words containing mostly unfamiliar letters ($p$ =.118). Results also indicated a significant difference in group performance in Block 3. However, this difference was for words that were made up of mostly unfamiliar letters ($p$ = .011). The LT-High group outperformed both the No-LT ($p < .05$) and the LT-Low groups ($p < .01$).

**Figure 5.3**

*Mean Accuracy in Word Learning by Block and Word Type for Letter Training Groups (LT-High, LT-Low, No-LT)*






*Note.* FAM Words = mostly familiar letters; unFAM Words = mostly unfamiliar letters.
**p* < .05. ***p* < .01. ****p* < .001. Error bars show standard errors: Confidence interval 95%.

Accuracy for each group by block and word type is shown in Figure 5.3. The LT-High group became more accurate between Blocks 2 and 3 for both word types: FAM ($p$ =.043) and unFAM ($p$ =.001). The No-LT group exhibited a different learning pattern but also performed consistently, becoming more accurate between Blocks 1 and 2 for both FAM ($p$ <.001) and unFAM words ($p$ <.001). However, results for the low-performing letter-training group differed. When words contained mostly familiar letters, the LT-Low were more accurate in Block 3 after having seen the word twice ($p < .05$). When words were made up of mostly unfamiliar letters, they were more accurate in Block 2 after only one exposure to the word ($p < .05$). No other pair-wise comparisons were significant and no other two or three-way interactions were found (*p*'s > .05).

**Table 5.6**

*Two and Three-Way ANOVA Statistics for Immediate and Delayed Word Type Data*

Immediate

| Measure | ANOVA | | | | | | |
|---|---|---|---|---|---|---|---|
| | | By subject | | | By item | | |
| | Effect | $F_1$ ratio | *df* | η2 | $F_2$ ratio | *df* | η2 |
| Word recognition - accuracy | WT | 7.979** | 1,39 | .170 | 4.045 | 1,39 | .094 |
| | WT × G | 1.214 | 2,39 | .059 | .501 | 2,39 | .025 |
| | WT × B | 1.336 | 2,78 | .033 | 1.758 | 2,78 | .043 |
| | B × G | 4.100** | 4,78 | .174 | 4.785** | 4,78 | .197 |
| | WT × B × G | .892 | 4,78 | .044 | 1.31 | 4,78 | .063 |
| Word recognition - response time | WT | 3.386 | 1,39 | .080 | 1.946 | 1,39 | .048 |
| | B | 44.510*** | 1.532,59.734 | .533 | 106.526*** | 1.615,62.991 | .732 |
| | G | .710 | 2,39 | .035 | 4.879* | 2,39 | .200 |
| | WT × G | 2.571 | 2,39 | .116 | 1.343 | 2,39 | .064 |
| | WT × B | .120 | 2,78 | .003 | .285 | 2,78 | .007 |
| | B × G | .712 | 4,78 | .035 | .792 | 4,78 | .039 |
| | WT × B × G | 2.112 | 4,78 | .098 | 1.107 | 4,78 | .054 |
| Recall - accuracy | WT | 5.516* | 1,39 | .124 | 2.380 | 1,39 | .058 |
| | G | 9.787*** | 2,39 | .334 | 12.024*** | 2,39 | .381 |
| | WT × G | 1.055 | 2,39 | .051 | .165 | 2.39 | .008 |

Delayed

| Measure | ANOVA | | | | | | |
|---|---|---|---|---|---|---|---|
| | | By subject | | | By item | | |
| | Effect | $F_1$ ratio | *df* | η2 | $F_2$ ratio | *df* | η2 |
| Word recognition - accuracy | WT | 3.337 | 1,38 | .081 | 1.836 | 1,39 | .045 |
| | G | 4.953* | 2,38 | .207 | 3.239 | 2,39 | .142 |
| | T | 7.228* | 1,38 | .160 | .514 | 1,39 | .013 |
| | WT × G | .407 | 2,38 | .021 | .190 | 2,39 | .010 |
| | WT × T | .213 | 1,38 | .006 | .291 | 1,39 | .007 |
| | G × T | .171 | 2,38 | .009 | 2.002 | 2,39 | .093 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | WT × G × T | .777 | 2,38 | .039 | 1.075 | 2,39 | .052 |
| Word recognition - response time | WT | 11.482** | 1,38 | .232 | .594 | 1,39 | .015 |
| | G | .515 | 2,38 | .026 | 2.089 | 2,39 | .097 |
| | T | 1.738 | 1,38 | .044 | 2.898 | 1,39 | .069 |
| | WT × G | .120 | 2,38 | .006 | .274 | 2,39 | .014 |
| | WT × T | .153 | 1,38 | .004 | .031 | 1,39 | .001 |
| | G × T | .007 | 2,38 | .000 | 1.126 | 2,39 | .055 |
| | WT × G × T | .557 | 2,38 | .028 | .054 | 2,39 | .003 |
| Recall - accuracy | WT | 8.077** | 1,38 | .175 | 3.438 | 1,39 | .081 |
| | G | 9.939*** | 2,38 | .343 | 11.866*** | 1,39 | .378 |
| | T | 7.135* | 1,38 | .158 | 10.967** | 1,39 | .219 |
| | WT × G | .077 | 2,38 | .004 | .001 | 2,39 | .081 |
| | WT × T | .920 | 1,38 | .024 | .542 | 1,39 | .014 |
| | G × T | .677 | 2,38 | .034 | .321 | 2,39 | .016 |
| | WT × G × T | 2.780 | 2,38 | .128 | 2.651 | 2,39 | .120 |

*Note.* ANOVA = analysis of variance; WT = word type; G = group; B = Block; T = time-point.
**p* < .05. ***p* < .01. ****p* < .001.

*Word Type (i.e., Number of Taught Letters) - Immediate Word Recognition Response Times*

A 3x2x3 three-way ANOVA evaluated response time with group (LT-High, LT-Low, No-LT) as a between-subject factor and word type (FAM, unFAM) and block (Block 1, Block 2, Block 3) as within-subject factors. The assumption of homogeneity of variances, as assessed by Levene's test, was met for all conditions ($p > .05$), with the exception of FAM words in Block 3 ($p = .037$). Homogeneity of covariances was not present, as assessed by Box’s test of equality of covariance matrices ($p < .001$). Mauchly's test of sphericity indicated that the assumption of sphericity was met for the block and word type interaction, $\chi2(2) = 4.243$, $p = .120$, but not for the main effect of block, $\chi2(2) = 13.870$, $p < .001$, thus, the Greenhouse-Geisser score is reported. As expected, there was a main effect of block ($p < .001$). Participants were significantly faster to respond to words with each subsequent presentation: FAM Blocks 1 and 2 ($p < .001$), FAM Blocks 2 and 3 ($p < .01$); unFAM Blocks 1 and 2 ($p < .01$), unFAM Blocks 2 and 3 ($p < .01$). There was also a main effect of group by item ($p = .013$), though it failed to meet significance by participant ($p = .498$). Words were responded to more slowly by the LT-High group (Block 1: $M = 2670.27$, $SD = 828.91$; Block 2: $M = 2228.72$, $SD = 704.66$; Block 3: $M = 1791.23$, $SD = 443.09$) compared to the LT-Low group (Block 1: $M = 2147.09$, $SD = 975.02$; Block 2: $M = 1725.39$, $SD = 415.32$; Block 3: $M = 1660.29$, $SD = 591.73$) when they were made up of unfamiliar graphemes ($p > .05$). No other main effects, two-way interactions, or three-way interactions were significant ($p$'s > .05).

*Word Type (i.e., Number of Taught Letters) - Immediate Recall Accuracy*

A 3x2 two-way ANOVA was carried out for accuracy in the recall task with group (LT-High, LT-Low, No-LT) as a between-subject factor and word type (FAM, unFAM) as a within-subject factor. Shapiro-Wilk's test of normality indicated data was normally distributed for all cases ($p$'s > .05). There was homogeneity of variances ($p > .05$), as assessed by Levene's test; Box's test showed homogeneity of covariances ($p > .05$). A main effect of word type was found ($p = .024$), though it failed to reach significance by item ($p = .131$). Participants were more accurate on words with mostly familiar graphemes ($p < .05$). There was also a main effect of group ($p < .001$). The LT-High group significantly outperformed the No-LT ($p < .01$) and LT-Low ($p < .01$) groups.

*Word Type (i.e., Number of Taught Letters) - Delayed Word Recognition Accuracy*

A 3x2x2 three-way ANOVA evaluated accuracy with group (LT-High, LT-Low, No-LT) as a between-subject factor and word type (FAM, unFAM) and time-point (Immediate-Block 3, Delayed-Block 4) as within-subject factors. The assumption of homogeneity of variances was met, as assessed by Levene's test for equality of variances ($p > .05$), with the exception of the immediate ($p = .025$) and delayed ($p = .038$) by item data for words made up of mostly familiar letters. There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices by subject ($p = .461$), but not by item ($p = .024$). A main effect of time-point was found ($p = .011$), though it failed to reach significance by item ($p = .478$). Participants were more accurate on the delayed measure (Block 4) for words with mostly familiar graphemes. A significant main effect was also found for group by participants ($p = .012$), though it was only approaching significance in the by item data ($p = .050$). The LT-High Group outperformed the No-LT ($p = .033$) and LT-Low ($p = .013$) groups when words contained mostly unfamiliar graphemes. No significant three-way or two-way interactions were found (*p's* > .05) and word type was not significant ($p = .076$).

*Word Type (i.e., Number of Taught Letters) - Delayed Word Recognition Response Times*

A 3x2x2 three-way ANOVA evaluated response time with group (LT-High, LT-Low, No-LT) as a between-subject factor and word type (FAM, unFAM) and time-point (Immediate-Block 3, Delayed-Block 4) as within-subject factors. The assumption of homogeneity of variances was met, as assessed by Levene's test for equality of variances ($p > .05$). Homogeneity of covariances, as assessed by Box's test of equality of covariance matrices was found by item ($p = .920$), but not by subject ($p = .003$). There was a main effect of word type by participants ($p = .002$), however, it failed to reach significance by item ($p = .446$). Participants were significantly faster ($p < .05$) to respond to words with mostly familiar letters. There was no significant three-way interaction by subject ($p = .578$) or by item ($p = .948$) and no other two-way interactions or main effects were significant ($p$'s $> .05$).

*Word Type (i.e., Number of Taught Letters) - Delayed Recall Accuracy*

A 3x2x2 three-way ANOVA evaluated accuracy with group (LT-High, LT-Low, No-LT) as a between-subject factor and word type (FAM, unFAM) and time-point (Immediate, Delayed) as within-subject factors. The assumption of homogeneity of variances was met, as assessed by Levene's test for equality of variances ($p > .05$) and there was homogeneity of covariances, as assessed by Box's test ($p > .05$). The analysis revealed a main effect of word type by subjects ($p = .007$), though it did not reach significance by items ($p = .071$). Participants were significantly more accurate ($p < .05$) on words with mostly familiar letters. There was also a significant effect of time-point ($p = .011$). Participants were more accurate ($p < .05$) on the delayed measure. Additionally, a significant difference in group performance was found ($p < .001$). The LT-High group outperformed the No-LT ($p < .01$) and LT-Low ($p < .01$) groups. No significant two or three-way interactions were found ($p$'s $> .05$).

*Additional Analyses*

Correlations between word accuracy and imageability, and word accuracy and image complexity, were explored but no significant effects were found. In the LT-High and LT-Low groups, correlations between placement of a familiar grapheme in the word-ending or word-initial positions for FAM items were also explored. It was thought that participants may have been applying first language (L1) strategies, despite the training slides on the direction in which Arabic is read. However, no significant effects for familiar grapheme position were found.

Lastly, participant data for number of languages learned and having learned a language with a different script were explored, but no correlations with participant accuracy or response times were found.

## Discussion

### *Overview*

The aim of this study was to assess the impact of letter training for absolute beginners in Arabic, in a multi-modal word learning task where an image and the pronunciation of a word was presented along with its written form in the Arabic script. A number of observations can be made regarding the results.

First, a learning benefit was clearly found for a high-performing subset of the letter-training group (LT-High). They outperformed the no training group (LT-No) on all immediate measures of learning. They also enjoyed an early learning benefit, performing above chance in the initial learning block and significantly outperforming the no letter training group. This indicates that participants who were successful in a letter training session were able to make use of the written input to learn Arabic words. Response times were not significantly different between participant groups in any of the blocks, indicating that either the letter-trained participants were not expending additional effort decoding the words or that the measure was too insensitive to detect it.

Secondly, the LT-High group outperformed the LT-No group on delayed measures of learning, suggesting a link between letter training and retention of learned material. Moreover, at the word level, there was evidence of increased accuracy and a slight processing advantage in the initial stages of the learning task for words with more familiar letters (Type-1 words) compared to those with no familiar letters (Type-4 words), or with fewer familiar letters. However, differences between word-types were no longer significant after two exposures to the target words.

Notably, a different learning pattern was observed between the groups. Participants who took part in a letter training (both LT-High and LT-Low participants) required three exposures to multi-modal stimuli to experience a significant gain in accuracy on the target words, compared to only two exposures for the LT-No group. These findings are discussed in greater detail in a later section.

Crucially, the results show that not all native English speakers who took part in the letter training were able to acquire the same amount of knowledge of the Arabic grapheme-phoneme associations. Put differently, a short 5-10 minutes of letter training did not allow all of the participants to achieve 100% accuracy in matching Arabic graphemes to phonemes, although all participants achieved fairly high levels of recognition (above 75%).

There were also significant differences between the accuracy of the high and low-performance subgroups on the immediate and delayed measures of learning, and the post-tests.

**Letter-training Results**

In the present study, just over half of the participants achieved 100% accuracy on the grapheme-phoneme test that followed the letter training. The low performance group had a mean accuracy of 85%. It may be that some of the participants needed more exposures to learn the grapheme-phoneme associations. For example, Bishop's (1964) participants achieved a mean accuracy of 99%. However, this was with 16 repetitions of the letters, compared to just 3 exposures in the current study. As there was no prior Arabic letter training research aside from

Bishop (1964) on which to base the current study's design, it would be a useful area for future research to explore the effect of exposures within the 3 to 16 range to determine a minimal number of additional exposures that can allow for a greater range of individual differences in working memory among participant groups.

Differences in the taught letters themselves, and the composition of the to-be-learned letter sets could also have resulted in variable outcomes. The current study's training included both the ر /r/ and the ز /z/, graphemes which differ only in the inclusion of a dot. Based on Laufer (1988) it was predicted this similarity might cause problems for some learners and results confirmed that the low-performance group was less accurate and slower in responding to these two letters in the grapheme-phoneme association test.

It is also possible that presenting auditory input in the form of the letter name, versus its corresponding phoneme, was problematic for the low-performance group. However, it would then have been expected to see poorer performance on letters such as the م /m/, where the taught letter name "meem" differed significantly from the associated phoneme, as compared to the ر /r/ where the letter name and phoneme match, yet the results showed no evidence of this.

*Performance of the Letter-training Low Group*

The low-performance group achieved the same accuracy (82%) as the no letter training control group in Block 3 of the word learning task. Therefore, it can be concluded that there was no effect for the letter training on word learning when participants scored below 100% on a test of grapheme-phoneme correspondence.

Interestingly, in looking at the low-performance group's accuracy in the initial block of the word learning task, there was evidence of a letter-training effect. In Block 1, participants were shown all items once in a match and once in a mismatch condition and instructed to guess at the meaning of words and use the feedback provided to learn them. None of the participants had any previous exposure to Arabic or to the to-be-learned items, nonetheless, participants in the high-performance group achieved 75% accuracy and the low-performance group 70% accuracy. This was compared to 63% accuracy for the no letter training group. Bitan and Karni (2003) found a similar result in their study in which only participants who had received explicit letter training performed at above chance on the first block of a translation task. This suggests participants in the low-performance group had acquired some knowledge of grapheme-phoneme mappings which they used to help them learn the words.

There is further evidence for this in examining the data on accuracy in Block 1 by word-type. For participants in the low-performance group, words where all of the letters were included in the training and in which only one letter was unfamiliar were responded to more accurately, at 74% and 71% respectively, compared to control words with no familiar letters and items with two unfamiliar letters, at 63% and 66% respectively. Nonetheless, no learning advantage was seen beyond Block 1.

Of note, the low-performance participants were the only group that experienced a significant decrease in accuracy in Session 2. This shows the learned material was more subject to attrition compared to the high-performance and no letter training groups. They were also the only group that experienced a significant gain in accuracy between the meaning-recall immediate and delayed post-tests. In other words, they were the only group that benefited from a fourth exposure to the words in Block 4.

Studies have shown that short-term memory for serial order and phonological awareness are both predictors for lexical learning in a second language (Majerus et al., 2018).

Consequently, individual differences could help explain why the low-performance group was less successful in both the letter training session and on measures of word-learning and retention. However, as this study did not include tests of working memory and phonological awareness, this is a question for future research.

An alternative explanation that is less likely but worth mentioning may be related to reading behavior. Wise, Yoncheva and McCandliss (2011) found participants who received a letter training in an artificial language divided into subgroups on a word-identification task, based on their preference for decoding or whole-word reading.

**Main Study Results**

*Performance of the Letter-training High Versus No Letter Training Group*

The high-performance group significantly outperformed the no letter training group in the word learning task and on a meaning-recall test, with a large effect size. As both groups were provided with the same visual, auditory, and written input, results suggest acquiring the phoneme-grapheme associations enabled absolute beginners to make use of the Arabic script, which helped them learn the words.

This differs from the findings of previous research where participants did not take part in a letter training and either received no benefit or a negative effect from exposure to the unfamiliar script (Mathieu, 2016; Showalter & Hayes-Harb, 2015). Nonetheless, it is important to consider that the current study also used a script manipulation in which the Arabic text was presented in a disconnected form. This potentially gave all participants an advantage in visually parsing the script, which could facilitate decoding of individual letters, resulting in greater form knowledge of the target words and enhanced performance (McCandliss et al., 1997). If this were the case, it might have been expected that the no letter training group in the present study was more successful than the written input group in Showalter and Hayes-Harb (2015) who were given connected word-forms. However, a comparative review of the performance of the two groups did not support this.

In looking at results compared to previous studies on letter training, performance trends are less clear. Findings are in line with those of Bolger (2007) and Bitan and Booth (2012), in which participants who were trained at the letter level significantly out-performed control groups. However, they differ from Brennan and Booth (2015) and Bitan and Karni (2003). In Brennen and Booth (2015) both letter-trained and whole-word trained participants achieved 90% accuracy in a word-matching test, and in Bitan and Karni (2003), there was no difference in the amount or accuracy of translations provided by the letter-training and control groups. One possible reason for this disparity may be the no letter training groups in these studies were trained on whole words, thus they received additional exposure to target language script, which was not the case for the no letter training group in our study. Moreover, the quality of the whole-word training participants in those studies received may have impacted their performance. Research has shown that adults can extract letter-sound rules when they are presented with L2 word lists which are representative enough to facilitate learning (Taylor et al., 2011).

In looking at response time results, findings were consistent with some previous research in which no difference was observed among participant groups (Bolger, 2007). However, they differed from studies in which letter-training resulted in processing delays (Brennan & Booth, 2015). In Brennan and Booth (2015) letter-trained groups were significantly slower across all tasks compared to whole-word trained groups. This may be due to the study’s focus on reading behavior. Whole-word groups may have been more likely to engage top-down processing. In the present study, the no letter training group did not receive any whole-word training before beginning the learning task. Moreover, participants were also learning with images, thus responses in the word learning task were not entirely dependent upon the decoding of written input. This may account for the difference in findings.

*Performance on Delayed Measures*

Participants in the high-performance letter-training group outperformed the no letter training group in the delayed measure of word learning and on a delayed meaning-recall test. These finding provide support for a link between letter training and retention of newly learned word-forms, as suggested by Bolger (2007). However, in Bolger's study participants outperformed a control group on a series of three tests, including two delayed measures. The current study's results are less telling. In fact, in looking at learning patterns, neither the high-performance nor the no letter training group experienced any significant decrease in performance between the immediate and delayed post-tests. Thus, it is not possible to attribute greater performance on the delayed measure to a retention-based benefit that comes from letter training alone.

*Differences between Words with More or Less Familiar Letters*

The current study investigated whether or not the number of familiar letters a word contained would have an effect on the accuracy and response time by word for participants from the letter training groups. An advantage in accuracy for the low-performance letter-training group was found in Block 1, as previously discussed, and a processing advantage in which words with all familiar and one unfamiliar letter were responded to faster, at 1841.43 ms and 1823.50

ms, compared to words with two unfamiliar and no familiar letters, at 2166.56 ms and 2211.76 ms was found for the letter-training high-performance group in Block 2. However, these results stemmed from planned comparisons and no significant differences were revealed by an ANOVA. Moreover, no differences in the accuracy or response time by word-types were found in Blocks 3 or 4, or the immediate and delayed post-tests. This suggests that familiar graphemes in the input only benefited participants in the very initial stages of form-meaning learning. Nonetheless, the relatively small set of stimuli in each category $N = 7$, and relatively low number of participants in each letter-training group, $N = 11$ and $N = 10$ may have meant not all differences reached significance. Further research into the number of familiar Arabic letters needed for a word to enjoy a learning or processing advantage may thus be of interest for future studies to explore.

*Learning Patterns across Blocks in a Multi-modal Task*

In Session 1 of the current study, two different learning patterns emerged. Participants who had received a letter-training achieved a significant increase in accuracy between Blocks 2 and 3 of the word learning task. This might suggest learning occurred following three exposures to the target words. In contrast, the no letter training group had a significant increase in accuracy between Blocks 1 and 2. This would suggest they required 2 exposures to learn the words and made no further gains with a subsequent exposure. Thus, it could be concluded that participants who took part in a letter training session engaged more with the written input and required an additional exposure. This could be due to the greater amount of attention they gave to the written forms, which were competing with visual and auditory stimuli for processing resources.

## Chapter 6. Conclusion

As Vocik states in discussing users of languages with considerably distant writing systems, “while the need to assemble phonological forms from graphemes is critical for all orthographies, the skills required to carry out this process effectively vary according to the nature of the orthography,” (2011, p. 395). In other words, learning to extract grapheme-phoneme correspondences from Arabic words alone is not necessarily a straightforward process for English native speakers. This is clear from previous research where providing written forms in the Arabic script had no positive effect, and in some cases, even negatively impacted word and phonemic-contrast learning for absolute beginners.

Explicit letter training thus presents an opportunity for Arabic as a Foreign Language (AFL) learners to quickly gain a foothold in the Arabic script, while also developing their awareness of target language phonology. More importantly, as demonstrated by the previously presented study’s findings, letter-training can be linked to vocabulary learning benefits. Absolute beginners with no prior exposure to Arabic successfully learned eight letters and their corresponding sounds and outperformed a no letter training group on both immediate and delayed measures of word learning. Participants were taught 28 words and achieved 92% accuracy on a word-recognition measure. Moreover, they were able to recall meaning when presented with auditory and written forms with 62% accuracy after only three exposures to the target words, and they maintained knowledge of the learned forms when tested one week later. Furthermore, there was evidence that letter-training can lead to decoding and processing advantages at the very initial stages of word-learning, regardless of success rates in a letter-training session.

Clearly, more research is needed to explore optimal approaches to training for absolute beginners. Roughly half of the participants who took part in the five-minute letter training session fell into a low-performance subset and did not experience the same performance advantages in the word learning task as those who successfully acquired the taught grapheme-phoneme correspondences. It is important to probe optimal number of item exposure repetitions, division of to-be-learned letters into groups of visually similar or dissimilar sets, and optimal audio input to accompany letter forms. While outside the scope of this study, it would also be of interest to curriculum developers and teachers to further understand the impact of letter training with actual AFL learners in the classroom and AFL learners with different L1s to English, particularly when the L1 utilizes an orthographic system more similar to Arabic. Understanding optimal letter training session length, frequency and usage alongside vocabulary focused lessons is a key area for future research. Nonetheless, building on the work of Bishop (1964), this study represents an important and current step in developing our understanding of Arabic letter training for English speakers.

In a study of Second Language (L2) adult vocabulary learning in AFL, Tseng, Doppelt and Tokowicz (2018) provided absolute beginners with either no written Arabic word forms or transliterations, justifying the use of a Latin script because it would support engagement with the spoken forms and "improve L2 Arabic vocabulary learning," (2018, p. 143). Yet for most institutions that offer Arabic Language courses, teaching with transliterations is not an option. Learners are introduced to the Arabic alphabet early on in beginner level courses, and vocabulary words are presented in the Arabic script in introductory materials. This is despite the fact that the Arabic script presents an initial learning challenge for students who may experience difficulties in letter recognition and formation, word recognition, spelling, and decoding as a result of both the lack of familiarity and the unique characteristics of Arabic orthography.

Given additional acknowledgement of the difficulties L1 English AFL learners face in acquiring Arabic vocabulary, it’s therefore crucial to produce more empirical research that informs our understanding of optimal letter-training approaches and vocabulary presentation for absolute beginners.

This leads to an obvious limitation of the letter training study in its application to an AFL learning context. A script manipulation in which the Arabic script was presented using isolated and spaced letter forms was used. It would, therefore, be important to replicate the findings of this research using authentic Arabic text presentation so as to be able to draw conclusions that could influence curriculum and materials development for Arabic as Foreign Language programs.

Lastly, while both studies in this thesis provide empirical support for the positive effects of multi-modal stimulus presentation in early AFL word-learning, it would be of interest to look at multi-modal learning when different constraints are applied to the to be learned stimulus set. Yet, as Mohamed (2016) states in discussing his own study of task-based incidental learning, “The importance of this study lies in its exploration of tasks and materials in Arabic language instruction as a recently evolving field of study … no considerable research has been done in Arabic vocabulary learning,” (pp. 43-44). It is clear more research into AFL vocabulary learning is needed and thus the greater aim of this thesis is to contribute to the growing body of AFL literature and support other researchers working in the field.

## References


AbdelRaouf, A., Higgins, C. A., Pridmore, T., & Khalil, M. (2010). Building a multi- modal Arabic corpus (MMAC). *International Journal on Document Analysis and Recognition*, *13*(4), 285-302. https://doi.org/10.1007/s10032-010-0128-2

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing*, *14*(1-2), 39-59.

Abu-Rabia, S. (2002). Reading in a root-based-morphology language: The case of Arabic. *Journal of Research in Reading*, *25*(3), 299-309.

Abuleil, S., & Alsamara, K. (2004). New technique to support Arabic noun morphology: Arabic Noun Classifier System (ANCS). *International Journal of Computer Processing of Oriental Languages*, *17*(02), 97-120. https://doi.org/10.1142/S0219427904001048

Adwan-Mansour, J., & Bitan, T. (2017). The effect of stimulus variability on learning and generalization of reading in a novel script. *Journal of Speech and Language Hearing Research*, *60*(10), 2840-2851.

Al Batal, M. (2006). Facing the crisis: Teaching and learning Arabic in the United States in a post-September 11 era. *Association of Departments of Foreign Languages Bulletin*, *37*(2-3), 39-46.

Al Mahmoud, M. S. (2013). Discrimination of Arabic contrasts by American learners. *Studies in Second Language Learning and Teaching*, *3*(2), 261-292.

Al-Batal, M. (1992). *Diglossia proficiency: The need for an alternative approach to teaching*. Wayne State University Press.

Al-Sulaiti, & Atwell. (2006). The design of a corpus of Contemporary Arabic. *International Journal of Corpus Linguistics*, *11*(2), 135-171.

Al-Sulaiti, L. (2010). *Arabic Corpora*. Retrieved February from http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm

Alduais, A. M. S., & Almukhaizeem, Y. S. (2015). Examining the effect of interference on short-term memory recall of Arabic abstract and concrete words using free, cued, and serial recall paradigms. *Advances in Language and Literary Studies*, *6*(6), 7-24.

Allen, D. (2013). *Cross-linguistic similarity in Japanese-English bilingual processing and representation* University of Nottingham]. Nottingham, UK.

Alyahya, R. S. W., & Druks, J. (2016). The adaptation of the Object and Action Naming Battery into Saudi Arabic. *Aphasiology*, *30*(4), 463-482. https://doi.org/10.1080/02687038.2015.1070947

Al-Sughaiyer, I. A., & Al-Kharashi, I. A. (2004). Arabic morphological analysis techniques: A comprehensive survey. *Journal of the American Society for Information Science and Technology*, *55*(3), 189–213. https://doi.org/10.1002/asi.10368

Amery, H. (2014). *Usborne first thousand words in Arabic*. Usborne.

Anderson, P., & Suleiman, Y. (2009). Arabic on campus and beyond. *British Journal of Middle Eastern Studies*, *36*(1), 125–144. https://doi.org/10.1080/13530190902749655

Apfelbaum, K., Hazeltine, E., & McMurray, B. (2012). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology*, *49*(7).

Asaad, H., & Eviatar, Z. (2013). The effects of orthographic complexity and diglossia on letter naming in Arabic: A developmental study. *Writing Systems Research*, *5*(2), 156–168. https://doi.org/10.1080/17586801.2013.862163

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). Academic Press.

Attia, M., Pecina, P., Toral, A., Tounsi, L., & Van Genabith, J. (2011). A lexical database for Modern Standard Arabic interoperable with a finite state morphological transducer. In C. Mahlow & M. Piotrowski (Eds.), *Systems and frameworks for computational morphology* (pp. 98–118). Springer. https://doi.org/10.1007/978-3-642-23138-4_7

Atwell, E., Al-Sulaiti, L., Al-Osaimi, S., & Abu-Shawar, B. (2004, April). A review of corpus analysis tools. JEP-TALN Arabic language processing, Fez, MA.

Baddeley, A. D. (2001). Is working memory still working? *The American Psychologist*, *56*(11), 851.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. Academic Press.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233–252. https://doi.org/10.1080/14640748408402157

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition*, *35*(5), 953.

Bassetti, B. (2009). Effects of adding interword spacing on Chinese
reading : A comparison of Chinese native readers and English

readers of Chinese as a second language *Applied Psycholinguistics*, *30*(4), 757–775.
Bassetti, B., Escudero, P., & Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, *36*(1), 1–6. https://doi.org/10.1017/S0142716414000393
Belnap, K. (1987). Whos taking Arabic and what on Earth for? A survey of students in Arabic language programs. *Al-Arabiyya*, *20*(1/2), 13.
Belnap, K. (2006). A profile of students of Arabic in U.S. universities. In K. Wahba, Taha, Z., & L. England (Eds.), *Handbook for Arabic language teaching professionals in the 21st century* (pp. 437–446). Routledge.
Best, C., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. https://doi.org/10.1037/0096-1523.14.3.345
Bishop, C. H. (1964). Transfer effects of word and letter training in reading. *Journal of Verbal Learning and Verbal Behavior*, *3*(3), 215–221. https://doi.org/10.1016/S0022-5371(64)80044-X
Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2013). Incidental acquisition of foreign language vocabulary through brief multi-modal exposure. *PLoS One*, *8*(4). https://doi.org/10.1371/journal.pone.0060912
Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014a). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, *64*(4), 855–877. https://doi.org/10.1111/lang.12085
Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014b). The role of verbal and pictorial information in multi-modal incidental acquisition of foreign language vocabulary. *The Quarterly Journal of Experimental Psychology*, 1–56. https://doi.org/10.1080/17470218.2014.979211
Bitan, T., & Booth, J. R. (2012). Offline improvement in learning to read a novel orthography depends on direct letter instruction. *Cognitive Science*, *36*(5), 896–918. https://doi.org/10.1111/j.1551-6709.2012.01234.x
Bitan, T., & Karni, A. (2003). Alphabetical knowledge from whole words training: effects of explicit instruction and implicit experience on learning script segmentation. *Cognitive Brain Research*, *16*(3), 323–337. https://doi.org/10.1016/S0926-6410(02)00301-4
Bolger, D. J. (2007). *The development of orthographic knowledge: A cognitive neuroscience investigation of reading skill* University of Pittsburgh]. Pittsburgh.

Boudelaa, S., Norris, D., Mahfoudhi, A., & Kinoshita, S. (2019). Transposed letter priming effects and allographic variation in Arabic: Insights from lexical decision and the same–different task. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(6), 729–757. https://doi.org/10.1037/xhp0000621

Boukadi, M., Zouaidi, C., & Wilson, M. A. (2016). Norms for name agreement, familiarity, subjective frequency, and imageability for 348 object names in Tunisian Arabic. *Behavior Research Methods*, *48*(2), 585. https://doi.org/10.3758/s13428-015-0602-3

Brennan, C., & Booth, J. R. (2015). Large grain instruction and phonological awareness skill influence rime sensitivity, processing speed, and early decoding skill in adult L2 learners. *Reading and Writing*, *28*(7), 917–938.

Brill, M. (1940). *The basic word list of the Arabic daily newspaper*. The University Publishing Company.

Brosh, H. (2015). Arabic spelling: Errors, perceptions, and strategies. *Foreign Language Annals*, *48*(4), 584–603. https://doi.org/10.1111/flan.12158

Brustad, K., Al-Batal, M., & Al-Tonsi, A. (2004). *Al-kitaab fii taallum al-arabiyya with DVDs: A textbook for beginning Arabic, part one*. Georgetown University Press.

Buckwalter, T. (2002). *Buckwalter Arabic Morphological Analyzer* (1.0 ed.). Linguistic Data Consortium (LDC).

Buckwalter, T., & Parkinson, D. (2011). *A frequency dictionary of Arabic: Core vocabulary for learners*. Routledge.

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge.

Cicerchia, M. (2010). *Arabic as a Foreign Language: Facilitating vocabulary acquisition through keyword strategies* University of Oxford]. Oxford, UK.

Cobb, T., & Horst, M. (2004). *Is there room for an Academic Word List in French?* Benjamins. https://doi.org/10.1075/lllt.10.04cob

Cook, V., & Bassetti, B. (2005). An introduction to researching second
language writing systems. In V. Cook & B. Bassetti (Eds.), *Second*
*language writing systems*. Multilingual Matters.

Curran, T., & Doyle, J. (2011). Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. *Journal of Cognitive Neuroscience*, *23*(5).

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*(2), 171–237.

Cárcamo, M. M. A., Cartes, R. A. C., Velásquez, N. E. E., & Larenas, C. H. D. (2015). The impact of multimodal instruction on the acquisition of vocabulary. *Trabalhos em Linguística Aplicada*, *55*(1).

Dai, J., Ibrahim, R., & Share, D. L. (2013). The influence of orthographic structure on printed word learning in Arabic. *Writing Systems Research*, 1–25. https://doi.org/10.1080/17586801.2013.827563

Daniels, P. (2013). The Arabic Writing system. In J. Owens (Ed.), *The Oxford handbook of Arabic Linguistics* (pp. 349–360). Oxford University Press.

Davidoff, J., & Masterson, J. (1996). The development of picture naming: Differences between verbs and nouns. *Journal of Neurolinguistics*, *9*(2), 69–83. https://doi.org/10.1016/0911-6044(96)00004-8

Desrochers, A., & Thompson, G. (2009). Subjective frequency and imageability ratings for 3,600 French nouns. *Behavior Research Methods*, *41*(2), 546–557. https://doi.org/10.3758/BRM.41.2.546

Detey, S., & Nespoulous, J. (2008). Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua*, *118*(1), 66–81. https://doi.org/10.1016/j.lingua.2007.04.003

Druks, J. (2002). Verbs and nouns – A review of the literature. *Journal of Neurolinguistics*, *15*(3–5), 289–315. https://doi.org/10.1016/S0911-6044(01)00029-X

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: New evidence from Dutch-French bilinguals. *Experimental Psychology*, *56*(3), 173. https://doi.org/10.1027/1618-3169.56.3.173

Duñabeitia, J. A. (2017). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 1. https://doi.org/10.1080/17470218.2017.1310261

Ebbinghaus, H. (1913). *Memory: a contribution to experimental psychology*. New York: Teachers College, Columbia University. https://doi.org/10.1037/10011-000

ECCE Guidelines for item writers. (2014). In C. M. L. Assessments (Ed.). Ann Arbor, Michigan.

Elkhafaifi, H. (2005a). Listening comprehension and anxiety in the Arabic language classroom. *Modern Language Journal*, *89*(2), 206–220. https://doi.org/10.1111/j.1540-4781.2005.00275.x

Elkhafaifi, H. (2005b). The effect of prelistening activities on listening comprehension in Arabic learners. *Foreign Language Annals*, *38*(4), 505–513. https://doi.org/10.1111/j.1944-9720.2005.tb02517.x

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Human Experimental Psychology*, *46*(3), 533-558. https://doi.org/10.1080/14640749308401062

Elsayyad, H., Everatt, J., Mortimore, T., & Haynes, C. (2017). The influence of working memory on reading comprehension in vowelized versus non-vowelized Arabic. *Reading and Writing: An Interdisciplinary Journal*, *30*(4), 871-886. https://doi.org/10.1007/s11145-016-9705-1

England, L. (2006). Methodology in Arabic language teacher education. In K. Wahba, Z. Taha, & L. England (Eds.), *Handbook for Arabic Language Teaching Professionals in the 21st Century* (pp. 419-436). Routledge.

Erce, A. V. (2008). *The effects of animation and cueing in the acquisition of Spanish vocabulary* Auburn Unviersity]. Auburn, Alabama.

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, *36*(1), 7-22. https://doi.org/10.1017/S014271641400040X

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345-360. https://doi.org/10.1016/j.wocn.2007.11.002

Eviatar, Z., Ganayim, D., & Ibrahim, R. (2004). Orthography and the hemispheres: Visual and linguistic aspects of letter processing. *Neuropysychology*, *18*(1), 174-184.

Familiar, L. (2021). A Frequency Dictionary of Contemporary Arabic Fiction: Core Vocabulary for Learners and Material Developers. In *Routledge Frequency Dictionaries Series*.

Fattah, M. A., & Ren, F. (2008). English-Arabic proper-noun transliteration-pairs creation. *Journal of the American Society for Information Science and Technology*, *59*(10), 1675-1687. https://doi.org/10.1002/asi.20877

Ferguson, C. A. (1959). Diglossia. *Word*, *15*(2), 325-340. https://doi.org/10.1080/00437956.1959.11659702

Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: A case study. *Language Learning Journal*, *36*(2), 239-248. https://doi.org/10.1080/09571730802390759

Fragman, A., & Russak, S. (2009, September). *A qualitative analysis of spelling errors in Arabic as a Foreign Language among native Hebrew speaking students* Arabele2009: Teaching and learning the Arabic language, FSI. (2020). *Foreign Language Training*. US Department of State. Retrieved January 24 from https://www.state.gov/foreign-language-training/

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(5), 1122-1139.

Golonka, E., Bowles, A., Silbert, N., Kramasz, D., Blake, C., & Buckwalter, T. (2015). The role of context and cognitive effort in vocabulary learning: A study of intermediate-level learners of Arabic. *Modern Language Journal*, *99*(1), 19-39. https://doi.org/10.1111/modl.12191

Habash, N., AbuOdeh, M., Taji, D., Faraj, R., El Gizuli, J., & Kallas, O. (June 2022). Camel treebank: An open multi-genre Arabic dependency treebank. 13th Conference on Language Resources and Evaluation (LREC 2022), Marseille, France.

Hamdi, R., Ghazali, S., & Barkat-Defradas, M. (2005, September). Syllable structure in spoken Arabic: A comparative investigation. Interspeech, Lisbon, PT.

Han, J.-I., Jeon, M., & Oh, S. (2017). Examining the temporal development of phonetic and lexical learning in second language. *Psychological Reports*, *120*(5), 785-804. https://doi.org/10.1177/0033294117707946

Han, J.-I., & Oh, S. (2018). The role of phonetic similarity and orthographic information in asymmetrical lexical encoding in second language. *Journal of Psycholinguistic Research*, 1-19. https://doi.org/10.1007/s10936-018-9574-7

Hansen, G. F. (2008). Visual word recognition in Arabic: Towards a language specific reading model. *University of Southern Denmark Centre for Contemporary Middle East Studies Working Paper Series*, *13*, 1-74.

Hansen, G. F. (2010). Word recognition in Arabic as a foreign language. *Modern Language Journal*, *94*(4), 567-581. https://doi.org/10.1111/j.1540-4781.2010.01094.x

Hary, B. (1996). *The importance of the language continuum in Arabic multiglossia*. American University in Cairo Press.

Hayes-Harb, R., & Cheng, H. W. (2016). The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word forms by native English speakers. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00785

Hayes-Harb, R., & Durham, K. (2016). Native English speakers' perception of Arabic emphatic consonants and the influence of vowel context. *Foreign Language Annals*, *49*(3), 557-572. https://doi.org/10.1111/flan.12217

Heddaya, A. (1985). *Qalam: A convention for morphological, Arabic-Latin-Arabic transliteration*. Retrieved February 2017 from http://langs.eserver.org/qalam.txt

Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological–syntactic disambiguation of homographic verbs and visual crowding effects. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(2), 494–507. https://doi.org/10.1037/xhp0000032

Hockley, W. E., & Bancroft, T. (2011). Extensions of the picture superiority effect in associative recognition. *Canadian Journal of Experimental Psychology /Revue canadienne de psychologie expérimentale*, *65*(4), 236–244. https://doi.org/10.1037/a0023796

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, *21*(2), 181–193.

Ibrahim, R. (2006). Morpho-phonemic similarity within and between languages: A factor to be considered in processing Arabic and Hebrew. *Reading and Writing*, *19*(6), 563–586. https://doi.org/10.1007/s11145-006-9009-y

Ibrahim, R., & Aharon-Peretz, J. (2005). Is literary Arabic a second language for native Arab speakers?: Evidence from semantic priming study. *Journal of Psycholinguistic Research*, *34*(1), 51–70. https://doi.org/10.1007/s10936-005-3631-8

Ibrahim, R., Eviatar, Z., & Aharon-Peretz, J. (2002). The characteristics of Arabic orthography slow its processing. *Neuropsychology*, *16*(3), 322–326. https://doi.org/10.1037//0894-4105.16.3.322

*International Corpus of Arabic*. (2013). Bibliotheca Alexandrina.

Khaldieh, S. (1996). Word recognition of Arabic as a foreign language by American learners: The role of phonology and script. *Al-Arabiyya*, *29*, 129–147.

Khoury, G. (2008). *Vocabulary acquisition in Arabic as a foreign language: The root and pattern strategy* University of Boston]. Boston, MA.

Khrisat, A. A., & Mohamad, M. S. (2014). Languages borrowings: The role of the borrowed and arabized words in enriching Arabic language. *American Journal of Humanities and Social Sciences*, *2*(2), 133–142. https://doi.org/10.11634/232907811402533

Khwaileh, T., Body, R., & Herbert, R. (2014). A normative database and determinants of lexical retrieval for 186 Arabic nouns: Effects of psycholinguistic and morpho-syntactic variables on naming latency. *Journal of Psycholinguistic Research*, *43*(6), 749–769. https://doi.org/10.1007/s10936-013-9277-z

Kuntz, P. (1996). Students of Arabic: Beliefs about foreign language learning. *Al-Arabiyya*(29), 153–176.

Kuntz, P. (2000). Beliefs about language learning: Students and their teachers at Arabic programs abroad. *African Issues*, *28*(1), 69–75.

Lababidi, Z. (2016). *The L2 perceptual mapping of Arabic and English consonants by American English learners* (Publication Number Thesis and Dissertations Paper 1288) University of Wisconsin-Milwaukee].

Lababidi, Z., & Park, H. (2014, September). Cross-language perceptual mapping between Arabic and English consonants. 19th Mid-continental Phonetics and Phonology Conference, Madison, WI.

Landau, J. M. (1959). *A word count of modern Arabic prose*. ACLS.

Laufer, B. (1988). The concept of "synforms" (similar lexical forms) in L2 vocabulary acquisition. *Language and Education*, *2*(2), 113-132.

Laufer, B. (1997). Whats in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge University Press.

Laufer, B. (2005). Focus on form in second language vocabulary acquisition. In S. Foster-Cohen (Ed.), *EUROSLA Yearbook 5* (pp. 223–250). John Benjamins

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1-26.

Lawson, D. R. (2010). An assessment of Arabic transliteration systems. *Technical Services Quarterly*, *27*(2), 164-177. https://doi.org/10.1080/07317130903547493

Liu, Y.-T., & Todd, A. G. (2016). Implementation of assisted repeated reading techniques for the incidental acquisition of novel foreign vocabulary. *Language Teaching Research*, *20*(1), 53-74. https://doi.org/10.1177/1362168814559802

Lotto, L., & de Groot, M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*(1), 31-69. https://doi.org/10.1111/1467-9922.00032

Lusin, N., & Looney, D. (2018). *Enrollments in Languages Other Than English in United States Institutions of Higher Education, Summer 2016 and Fall 2016: Preliminary Report*. M. L. A. o. America.

Macaro, E. (2001). *Learner strategies in second and foreign language classrooms*. Continuum.

Maghrabi, A. (1997). *The roles of psycholinguistic constraints and typological influence in the acquisition of pronominal copies in relativization by Arabic and English learners* Georgetown University]. Washington, DC.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. (2018). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, *107*(2), 395–419.

Masterson, J., & Druks, J. (1998). Description of a set of 164 nouns and 102 verbs matched for printed word frequency, familiarity and age-of-acquisition. *Journal of Neurolinguistics*, *11*(4), 331–354.

Masterson, J., Druks, J., & Gallienne, D. (2008). Object and action picture naming in three- and five- year-old children. *Journal of Child Language*, *35*(2), 373–402. https://doi.org/10.1017/S0305000907008549

Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, *32*(2), 145–170. https://doi.org/10.1177/0267658315601882

McCandliss, B., Schneider, W., & Smith, T. (1997). *Learning to read new visual symbols as integrated wholes or component parts* 38th Annual Meeting of the Psychonomic Society, Philadelphia, PA.

McLaughlin, B. (1987). *Theories of second language learning*. Arnold.

McLeod, B., & McLaughlin, B. (1986). Restructuring of automaticity? Reading in a second language. *Language Learning*.

Mohamed, A. (2016). Task-based incidental vocabulary learning in L2 Arabic: The role of proficiency and task performance. *Journal of the National Council of Less Commonly Taught Languages*, *18*, 121–157. https://doi.org/10.10117/issn.1930-9031.2015.03.08

Morin, R. (2003). Derivational morphological analysis as a strategy for vocabulary acquisition in Spanish. *Modern Language Journal*, *87*(2), 200–221.

Moser, J. (2013). *Bringing the lexical approach to TAFL: Evaluating the primary lexicon in part one of the Al-Kitaab Fii Taallum Al-Arabiyya Arabic as a Foreign Language textbook series* University of Arizona]. Tuscon, AZ.

Nassaji, H. (2004). Input modality and remembering name-referent associations in vocabulary learning. *Canadian Journal of Applied Linguistics*, *7*(1), 39–55.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle & Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nevat, M., Khateb, A., & Prior, A. (2014). When first language is not first: An functional magnetic resonance imaging investigation of the neural basis of diglossia in Arabic. *European Journal of Neuroscience*, *40*(9), 3387–3395. https://doi.org/10.1111/ejn.12673

Olshtain, E. (1987). The acquisition of new word formation processes in second language acquisition. *Studies in Second Language Acquisition*, *9*(2), 221–231. https://doi.org/10.1017/S0272263100000486

Omari, S. E. (2015). The effect of computer-assisted language learning on improving Arabic as a foreign language (AFL) in higher education in the United States. *Procedia - Social and Behavioral Sciences*, *192*, 621–628. https://doi.org/10.1016/j.sbspro.2015.06.109

Oweini, A., & Hazoury, K. (2010). Towards a sight word list in Arabic. *International Review of Education*, *56*(4), 457–478. https://doi.org/10.1007/s11159-010-9170-z

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart and Winston.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short- term memory and foreign- language vocabulary learning. *Journal of Memory and Language*, *30*(3), 331–347. https://doi.org/10.1016/0749-596X(91)90040-Q

Perea, M., Mallouh, R. A., & Carreiras, M. (2014). Are root letters compulsory for lexical access in Semitic languages? The case of masked form-priming in Arabic. *Cognition*, *132*(3), 491–500. https://doi.org/10.1016/j.cognition.2014.05.008

Ramezanzadeh, A. M. (2016). *One size fits all? An analysis of heritage and non-heritage language learner performance in GCSE Arabic*. British Council.

Ramezanzadeh, A. M. (2017, March). *Motivation and multiglossia: Developing a tool for examining Arabic learners motivation and engagement profiles* Arabic Language Teaching and Learning in UK Higher Education Conference, University of Leeds, UK.

Randall, M., & Meara, P. (1988). How Arabs read Roman letters. *Reading in a Foreign Language*, *4*(2), 133–145.

Ravid, D. (2005). A developmental perspective on root perception in Hebrew and Palestinian Arabic. In J. Shimron (Ed.), *Language processing and acquisition in languages of semitic, root- based, morphology* (pp. 293–319). J. Benjamins.

Ricks, R. S. (2015). *The development of frequency-based assessments of vocabulary breadth and depth for L2 Arabic* Georgetown University]. Washington, DC.

Russak, S., & Fragman, A. (2014). Spelling development in Arabic as a Foreign Language among native Hebrew speaking pupils. *Reading and Writing*, *27*(2), 359–381. https://doi.org/10.1007/s11145-013-9448-1

Ryding, K. C. (1991). Proficiency despite diglossia: A new approach for Arabic. *Modern Language Journal*, *75*(2), 212–218.

Ryding, K. C. (2013). *Arabic: A linguistic introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139151016

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics*, *24*(3), 431–451. https://doi.org/10.1017/S0142716403000225

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951. https://doi.org/10.1111/lang.12077

Schupp, H. T., Stockburger, J., Codispoti, M., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2007). Selective visual attention to emotion. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(5), 1082. https://doi.org/10.1523/JNEUROSCI.3223-06.2007

Schweickert, R., Guentert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, *1*(1), 74–77.

Shintani, N. (2013). The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of L2 vocabulary by young beginning-level learners. *Tesol Quarterly*, *47*(1), 36–62. https://doi.org/10.1002/tesq.54

Showalter, C. E., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, *36*(1), 23–42. https://doi.org/10.1017/S0142716414000411

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Sommers, M., & Barcroft, J. (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Applied Psycholinguistics*, *32*(2), 417–434.

Stevens, P. B. (1996). *Reflections on acquiring Arabic as a Second Language: One linguists perspective as a learner*. American University in Cairo Press.

Stowasser, B. F. (1981). Semantic analysis and the teaching of Arabic vocabulary. *Al-Arabiyya*, *14*, 5–10.

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278. https://doi.org/10.1016/j.neuroimage.2013.08.023

Tamminem, J., Davis, M., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39.

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 60–76.

*The British National Corpus*. (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium.

Tibi, S., Edwards, A. A., Schatschneider, C., Lombardino, L. J., Kirby, J. R., & Salha, S. H. (2020). IRT analyses of Arabic letter knowledge in Kindergarten. *Reading & writing*. https://doi.org/10.1007/s11145-020-10086-6

Towler, M. (2017, March 28). *Trapped between MSA and everyday tasks: Learning Arabic at HEIs in England* Arabic Language Teaching & Learning in UK Higher Education Conference, University of Leeds, Leeds.

Towler, M. (2018). *Language mapping pilot project: Arabic language provision in the UK*. T. B. Academy. https://www.thebritishacademy.ac.uk/

Tseng, A. M., Doppelt, M. C., & Tokowicz, N. (2018). The effects of transliterations, thematic organization, and working memory on adult L2 vocabulary learning. *Journal of Second Language Studies*, *1*(1), 141–165.

UCML-AULC. (2018). *Survey of Institution-Wide Language Provision in universities in the UK (2017-2018)*. http://www.aulc.org/wp-content/uploads/2018/04/AULC_UCML_2017-2018.pdf

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Van Mol, M. (2006). Arabic receptive language teaching: A new CALL approach. In K. M. Wahba, Z. Taha, & L. England (Eds.), *Handbook for Arabic language teaching professionals in the 21st century* (pp. 305–314). Laurence Erlbaum Publishers.

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, *35*(5), 938–952.

Vokic, G. (2011). When alphabets collide: Alphabetic first–language speakers' approach to speech production in an alphabetic second language. *Second Language Research*, *27*(3), 391–417.

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63–81. https://doi.org/10.1177/1362168806072463

Welles, E. B. (2002). Foreign language enrollments in the United States institutions of higher education. *ADFL Bulletin Fall*, *35*(2–3), 7–26.

Wesche, M., & Paribakht, T. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review–Revue Canadienne Des Langues Vivantes*, *53*(1), 13–40.

Wise, J., Yoncheva, Y., & McCandliss, B. (2011). Effects of preference and strategy on learning to read an artificial script. *Indiana University Undergraduate Journal of Cognitive Science*, *6*, 38–47.

Young, E. (2011). *Teaching and learning Arabic vocabulary: A learner–centered approach* University of Texas]. Austin, TX.

Zarei, G. R., & Khazaie, S. (2011). L2 vocabulary learning through multimodal representations. *Procedia – Social and Behavioral Sciences*, *15*, 369–375. https://doi.org/10.1016/j.sbspro.2011.03.104

APPENDIX 4.1 List of stimulus items and their characteristics used in the 'Multi-modal Arabic word learning in novice L1 English speakers' study

**Arabic Words (List 1)**

| List | Transliteration | Arabic | Letters | Syllables | Arabic-only phonemes |
|---|---|---|---|---|---|
| 1 | fiTr | فطر | 4 | 1 | 1 |
| 1 | Huwt | حوت | 4 | 1 | 1 |
| 1 | sahm | سهم | 4 | 1 | |
| 1 | mawz | موز | 4 | 1 | |
| 1 | taaj | تاج | 4 | 1 | |
| 1 | mishT | مشط | 5 | 1 | 1 |
| 1 | khass | خس | 5 | 1 | 1 |
| 1 | shams | شمس | 5 | 1 | |
| 1 | salla | سلة | 5 | 2 | |
| 1 | mi`Taf | معطف | 6 | 2 | 2 |
| 1 | mirsaa | مرساة | 6 | 2 | |
| 1 | arnab | أرنب | 6 | 2 | |
| 1 | ya`soob | يعسوب | 7 | 2 | 1 |
| 1 | Haafila | حافلة | 7 | 3 | 1 |
| 1 | miTraqa | مطرقة | 7 | 3 | 2 |
| 1 | jumjuma | جمجمة | 7 | 2 | |
| 1 | zujaaja | زجاجة | 7 | 3 | |
| 1 | kaneesa | كنيسة | 7 | 3 | |
| 1 | shajara | شجرة | 7 | 3 | |
| 1 | tannoora | تنورة | 8 | 3 | |

**Arabic Words (List 1 Continued)**

| List | Transliteration | Arabic | Letters | Syllables | Arabic-only phonemes |
|---|---|---|---|---|---|
| 1 | fa's | فأس | 4 | 1 | 1 |
| 1 | dhura | ذرة | 5 | 2 | 1 |
| 1 | najma | نجمة | 5 | 2 | |
| 1 | isTabl* | إسطبل | 6 | 2 | 1 |
| 1 | Hizaam | حزام | 6 | 2 | 1 |
| 1 | ghayma | غيمة | 6 | 2 | 1 |
| 1 | udhun | أذُن | 6 | 2 | 1 |
| 1 | miqaSS | مقص | 6 | 2 | 1 |
| 1 | Difda' | ضفدع | 6 | 2 | 2 |
| 1 | ibreeq | إبريق | 7 | 2 | 1 |
| 1 | booSala | بوصلة | 7 | 3 | 1 |
| 1 | miZalla | مظلة | 7 | 3 | 1 |
| 1 | Taa'ira | طائرة | 7 | 3 | 2 |
| 1 | areeka | أريكة | 7 | 3 | |
| 1 | thu`baan | ثعبان | 8 | 2 | 1 |
| 1 | tuffaaHa* | تفاحة | 8 | 3 | 2 |
| 1 | ghassaala | غسالة | 9 | 3 | 1 |
| 1 | thallaaja | ثلاجة | 9 | 3 | |
| 1 | mazhariya | مزهرية | 9 | 3 | |
| 1 | baazillaa' | بازلاء | 10 | 3 | |

*Removed from final analysis

**Arabic Words (List 2 Continued)**

| List | Transliteration | Arabic | Letters | Syllables | Arabic-only phonemes |
|---|---|---|---|---|---|
| 2 | Zarf | ظرف | 4 | 1 | 1 |
| 2 | `ayn | عين | 4 | 1 | 1 |
| 2 | dumya | دمية | 5 | 2 | |
| 2 | sham`a | شمعة | 6 | 2 | 1 |
| 2 | `ajala | عجلة | 6 | 3 | 1 |
| 2 | wi`aa' | وعاء | 6 | 2 | 2 |
| 2 | seeaaj | سياج | 6 | 2 | |
| 2 | Sabbaar | صبار | 7 | 2 | 1 |
| 2 | Hidhaa' | حذاء | 7 | 2 | 3 |
| 2 | khinzeer | خنزير | 8 | 2 | 1 |
| 2 | siHleeya | سحلية | 8 | 3 | 1 |
| 2 | sulHufaa | سلحفاة | 8 | 3 | 1 |
| 2 | muharrij | مهرج | 8 | 3 | |
| 2 | faraawla | فراولة | 8 | 3 | |
| 2 | naZZaarat | نظارة | 9 | 3 | 1 |
| 2 | barTamaan | برطمان | 9 | 3 | 1 |
| 2 | da`sooqah | دعسوقة | 9 | 3 | 2 |
| 2 | ghallaaya | غلاية | 10 | 3 | 1 |
| 2 | khunfusaa' | خنفساء | 10 | 3 | 2 |
| 2 | kushtubaan | كشتبان | 10 | 3 | |

**Arabic Words (List 2)**

| List | Transliteration | Arabic | Letters | Syllables | Arabic-only phonemes |
|---|---|---|---|---|---|
| 2 | zir | زر | 3 | 1 | |
| 2 | yad | يد | 3 | 1 | |
| 2 | qaws | قوس | 4 | 1 | 1 |
| 2 | bi'r | بئر | 4 | 1 | 1 |
| 2 | dubb | دب | 4 | 1 | |
| 2 | baab | باب | 4 | 1 | |
| 2 | anf | أنف | 4 | 1 | |
| 2 | deek | ديك | 4 | 1 | |
| 2 | kura | كرة | 4 | 2 | |
| 2 | baTTa | بطة | 5 | 2 | 1 |
| 2 | `alam | علم | 5 | 2 | 1 |
| 2 | jaras | جرس | 5 | 2 | |
| 2 | zahra | زهرة | 5 | 2 | |
| 2 | jabal | جبل | 5 | 2 | |
| 2 | timsaaH | تمساح | 7 | 2 | 1 |
| 2 | qilaada | قلادة | 7 | 3 | 1 |
| 2 | kurssee | كرسي | 7 | 2 | |
| 2 | sikkeen | سكين | 7 | 2 | |
| 2 | silsila | سلسلة | 7 | 3 | |
| 2 | sayyaara | سيارة | 8 | 3 | |