The effect of printed word attributes on Arabic reading

A thesis submitted to Lancaster University for the degree of
Doctor of Philosophy in the
Faculty of Science and Technology

September 2017

Ahmed Mohammed A Alhussein, B.Ed., M.Ed.

Department of Psychology

Declaration

This thesis is my own work and no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification at this or any other institute of learning.

Ahmed Alhussein 30th September 2017

Acknowledgements

I would like to give profound thanks to God for giving me the ability, strength and determination to complete my research journey. I am very grateful to the government of Saudi Arabia and to my university, Imam Muhammad Ibn Saud Islamic University, for their support throughout my Ph. D. study. I also owe a great deal to my supervisors, Dr. Robert Davies and Professor Gert Westermann. Dr. Davies in particular has been a constant source of advice and support, with an endless supply of patience. It is important to acknowledge the fantastic support I received from my wife and children during my research. I am extremely grateful for the moral support I have had from my family, especially my father. Finally, I wish to thank everyone else who has contributed to the completion of this work, especially the participants, without whom there would be no research.

Abstract

Printed Arabic texts usually contain no short vowels and therefore a single letter string can often be associated with two or more distinct pronunciations and meanings. The high level of homography is believed to present difficulties for the skilled reader. However, this is the first study to gather empirical evidence on what readers know about the different words that can be associated with each homograph. There are few studies of the effects of psycholinguistic variables on Arabic word naming and lexical decision. The present work therefore involved the creation of a database of 1,474 unvowelised letter strings, which was used to undertake four studies. The first study presented lists of unvowelised letter strings and asked participants to produce the one or more word forms (with short vowels) evoked by each target. Responses to 1,474 items were recorded from 445 adult speakers of Arabic. The number of different vowelised forms associated with each letter string and the percentage agreement were calculated. The second study collected subjective Age-of-Acquisition ratings from 89 different participants for the agreed vowelised form of each letter string. The third study asked 38 participants to produce pronunciation responses to 1,474 letter strings. Finally, 40 different participants were asked to produce lexical decisions to 1,352 letter strings and 1,352 matched non-word letter strings. Mixed-effects models showed that orthographic frequency, Age-of-Acquisition and name agreement influenced word naming, while lexical decision was not affected by name agreement. Findings indicate that lexical decision in Arabic requires recognition of a basic shared morphemic structure, whereas word naming requires identification of a unique phonological representation. It takes

longer to name a word when there are more possible pronunciations. The Age-of-Acquisition effect is consistent with a developmental theory of reading.

Table of Contents

Table of figures

Table of tables

Chapter 1 INTRODUCTION AND BACKGROUND

1.1 Introduction

Arabic is the fourth most common language in the world and is spoken by over 200 million native speakers from Asia to the Atlantic Ocean (Lewis, Simons, & Fennig, 2014; Paulston, Kiesling, & Rangel, 2012; Saiegh-Haddad & Henkin-Roitfarb, 2014). All native speakers acquire one of thirty or more varieties of spoken (colloquial) Arabic which differ in pronunciation, many items of vocabulary and word structure (Saiegh-Haddad & Henkin-Roitfarb, 2014). These varieties have no written form, in contrast to Modern Standard Arabic (MSA) which children learn when they start school. Worldwide, the Arabic script is the most commonly used phonemic script after Roman (Saiegh-Haddad & Henkin-Roitfarb, 2014). However, there has been relatively little research into Arabic linguistics (Owens, 2013). Yet the Arabic language has several characteristics of particular interest, including its morphology and orthography, such that the study of reading in Arabic may help to increase our understanding of the cognitive processes of reading in general. The present study will shed light on how printed text which consists mainly of consonants, omitting short vowels, is read. It specifically considers how native Arabic speakers access pronunciations in a highly ambiguous printed text when reading aloud and how they process visual recognition of isolated words.

This chapter introduces the rationale for this study, its nature and the research questions. It also sets out the structure for the whole thesis. The second chapter will familiarise the reader with the characteristics of the Arabic language. It will give the non-Arabic speaking reader a basic understanding of Arabic orthography. This will provide a platform for the reader to understand the later discussion that identifies homographs as the focus of the present study.

Chapter three reviews some of the key relevant literature, including processes and theories of visual word recognition and reading aloud, as well as critical findings from previous experimental research, while the fourth and fifth chapters describe the development of a database for use in the later stages of this research and other studies. Development of the database involved two studies. The first was a study of name agreement to determine which words were heterophonic homographs (different pronunciation, same spelling in printed text) and which of the possible alternative readings (pronunciations) of each homograph would be used as target words in the word naming experiment. Identification of the number of alternatives enabled measures of the extent of name agreement to be used as a variable in word naming and lexical decision experiments. The second study collected Age of Acquisition (AoA) ratings which were then used as a predictor variable in analyses of the data from a word naming and a lexical decision experiment. Models of predictive variables are presented in both studies, as is a discussion of the findings. By creating a database of Arabic words which intentionally includes heterophonic homographs, the present study aims to support experiments investigating the impact of homography on reading and the implications for theories of word recognition and reading

The sixth chapter deals with a word naming experiment involving 38 native speakers of Arabic, the hypotheses it tests, and discussion of the findings. The influence on word naming of name agreement, AoA, frequency, length and the interactions between these effects is examined using mixed effects regression modelling. Results are compared with findings from other studies and implications are discussed for theories about the process of reading aloud isolated words in Arabic. The seventh chapter follows a similar structure and describes and discusses the findings from a study using the lexical decision task. This second experiment involved

40 native speakers and, as for word naming, investigated the influence of name agreement, AoA, frequency, length and the interactions between these effects. Comparison of results with other studies and implications for understanding the processes of lexical decision and visual word recognition in Arabic conclude the chapter. Name agreement exerts a strong influence on word naming, while the effect of AoA is found to be strong in both word naming and lexical decision. The final chapter provides an overall discussion of the findings and their implications for theoretical account of how words are recognised and read aloud in Arabic text. The conclusion in chapter eight summarises the findings and proposes future directions for research.

1.2 Rationale for the study

The acquisition of language in humans is an area of considerable long-term interest, and the way in which words are recognized, sentences formed and reading occurs is a specific area that warrants extensive research. The ability to read is an extremely complex and unique skill and, despite a vast body of literature based chiefly on Anglocentric and western European areas of interest, relatively little is known or understood about how this skill is acquired in Semitic languages and other languages with different orthographies (Share, 2008).

Research focused on word recognition has been at the heart of theoretical accounts of language processing, memory and attention, not least because word analysis can be undertaken at different levels, such as letters, graphemes, morphemes and phonemes (Balota, Yap & Cortese, 2006). Moreover, word recognition research has increased understanding of the fundamental processes of pattern recognition (Balota et al., 2006). Patterns in Arabic orthography and word structure are different from those in English and other Indo-European languages, so that therefore their investigation may further increase understanding.

Much previous research in Arabic has focused on whether ambiguous printed words can be read in isolation or whether word recognition required a disambiguating sentence context. The ambiguity arises from a single unvowelised letter string mapping onto multiple vowelised forms. Some studies have asserted that context is necessary because it enables readers to identify or guess the vowels that are omitted in printed text (Abu-Rabia, 1997; Abu-Rabia, & Siegel, 1995). One study has indicated that native speakers can and do read isolated words successfully (Ibrahim, 2013). However, a resolution of previous findings is difficult because the results of the various studies are not directly comparable due to differences in purpose and critically, in stimulus types and subjects. An important step is therefore to identify a sufficient list of homographs for use in the present and future studies. It is necessary to establish which printed Arabic words are recognised by readers as homographs, that is, letter strings corresponding to multiple different words. There may be multiple dictionary entries for a single letter string, but readers may know only one of the possible alternatives. Very few studies have specifically investigated the effects of homography on reading in Arabic or have considered the implications for the cognitive processes involved or theories of visual word recognition and reading. The studies and experiments in this thesis offer a starting point.

In the search for explanations of the relationships between the phonological, semantic and written representations of words, it is asserted that the morphologies of languages such as Hebrew and Arabic are a rich resource because differences in word structure between Semitic and Indo-European languages may offer new insights into reading processes (Balota et al., 2006). Moreover, the development of the Aralex corpus by Boudelaa and Marslen-Wilson (2010) has opened new possibilities for psycholinguistic studies in Arabic.

The present study aims to use the rich resources of the Arabic language and exploit the existence of the Aralex corpus (Boudelaa & Marslen-Wilson, 2010) to explore how people read printed text which is highly homographic. Investigations into how people disambiguate homographs in different languages have identified that various factors influence how they do this. These factors include among others how often they encounter one particular meaning of the homograph rather than another in Hebrew (e.g. Bentin & Frost, 1987) and Arabic (Seraye, 2004), how many possible meanings there are and how closely the different meanings are related in English, with a focus on words that share pronunciation and spelling (e.g. Rodd, Gaskell & Marslen-Wilson, 2002), and reliance on their vocabulary knowledge and grammar rules in Hebrew (e.g. Share, 2008). The present thesis investigates the first two of these, asking the following research questions.

1.3 Research question

The broad research question is: How is reading in Arabic affected by a highly homographic orthography as it appears in printed text? This breaks down into specific questions: What is the effect of word frequency on visual recognition and reading aloud of isolated words in Arabic printed text? What is the effect of word length on visual recognition and aloud of isolated words in Arabic printed text? What is the effect of name agreement on visual recognition and reading aloud of isolated words in Arabic printed text? What is the effect of AoA on visual recognition and reading aloud of isolated words in Arabic printed text? What are the implications of observations of these effects for understanding the cognitive processes involved?

At this point, it is important to be clear about key definitions used in this thesis and to provide readers with an understanding of key features of the Arabic orthography and word structure.

1.4 Definitions of terms

Diacritics: These are symbols (not letters) placed above or below consonants in Arabic. They include:

FatHah (◌َ) which produces a sound similar to placing the letter *a* after *b* in the word *bat*

Dhammah (◌ُ) which produces a sound similar to placing the letter *u* after *p* in the word *put*

Kasrash (◌ِ) which produces a sound similar to placing the letter *i* after *b* in the word *bit*

Sukoon (◌ْ) which indicates that the consonant is not accompanied by a vowel sound

Shaddah (◌ّ) which indicates a doubling of the consonant (Ryding, 2005).

Arabic Short Vowels: These are the sounds produced by adding the abovementioned *fatHah*, *dhmammah* and *kasrah* to consonants (Ryding, 2005).

Heterophonic Homographs (in Arabic): This phenomenon is not found in Arabic when letter strings are printed accompanied by diacritics. It occurs when the letters appear without diacritics (short vowel signs), so that unvowelised letter strings (words printed without diacritics) are spelling-to-sound ambiguous (and thus necessarily spelling-to-meaning ambiguous) where there are multiple lexical entries corresponding to a letter string.

Homophonic Homographs (in Arabic): These are words with the same spelling and pronunciation, even when diacritics are present, but with different meanings, where two different meanings are associated with a single phonological and visual form (Peleg, Eviatar, Manevitz, & Hazan, 2007).

For clarity, this study focuses on *heterophonic homographs*, and will use either *heterophonic homograph* or *homophonic homograph* when necessary to clarify which one of them is under consideration, for example when referring to previous studies using those terms.

Nonwords: Words which conform to the orthographic and phonological patterns of a language but have no meaning in the language. These are also known as legal nonwords and pseudowords. Nonwords which do not conform to language patterns in this way are sometimes termed 'illegal nonwords' (Warren, 2012).

Chapter 2 THE ARABIC LANGUAGE: ORTHOGRAPHY AND WORD STRUCTURE

2.1 Introduction

There are a number of characteristics that contribute to the complexity of Arabic orthography. It is written from right to left, in a cursive script in which most letters are joined together but take a slightly different shape depending on their position in the word. When all the diacritics are present, there is almost a one-to-one correspondence of grapheme to phoneme, but in printed text this correspondence disappears. Since printed Arabic consist almost entirely of consonants, phonemic information from short vowels is almost completely absent. However, the word structure provides information about the meaning and syntax of the word. This chapter explains the key features of Arabic orthography and word structure that are necessary to understand the studies and experiments later in the thesis.

2.2 The Arabic alphabet

Arabic is written from right to left in a mainly consonantal alphabet (or ‘abjad’) with 28 basic graphemes plus one particular letter, hamza, which represents the glottal stop. Twenty five of the 28 basic graphemes are consonants, while three represent long vowels (أ، و، ي), as can be seen in Table 1.

Table 1 Arabic graphemes with corresponding IPA symbols

Arabic grapheme	IPA	Arabic grapheme	IPA
اِ	ɑ	ص	sˤ
ب	b	ض	dˤ
ت	t	ط	tˤ
ة	h	ظ	ðˤ
ث	θ	ع	ʕ
ج	dʒ	غ	ɣ
ح	ħ	ف	F
خ	x	ق	q
د	D	ك	K
ذ	ð	ل	l, ɫ
ر	r	م	m
ز	z	ن	n
س	S	ه	H
ش	ʃ	و	u, w

Source: IPA http://www.internationalphoneticalphabet.org

Three short vowels are written as diacritics instead of letters. Twenty letters belong to groups of letters in which the shape of the letter is the same but the letters are distinguished by the presence of dots (Breznitz, 2004). In other words, letters can be grouped according to their basic shapes and distinguished by the number of dots in, on, or below the letter, or whether the letter has dots or not, for example /ب،ت،ث/ج،ح،خ. These are set out together with their transliterations in Tables 2 and 3.

Table 2 Group of letters distinguished by dots: example 1

Name of letter in IPA	no dots	1 dot	2 dots	3 dots
bɑ		ب		
tɑ			ت	
θɑ				ث

Table 3 Group of letters distinguished by dots: example 2

Name of letter in IPA	no dots	1 dot	2 dots	3 dots
dʒi:m		ج		
ħɑ	ح			
xɑ		خ		

Comparison of Tables 2 and 3 illustrates how patterns of dots differ according to the basic shape of the letter group.

Another element of the perceptual load (Abdelhadi, Ibrahim, & Eviatar, 2011) resulting from the orthographical complexity of printed Arabic is the existence of positional variants of letters within the word. Most letters can be written in more than one form, depending on where the letter occurs in a word. The particular letter form to be used is determined by its position in a word, which is classified as initial, medial or final. An example for the letter Ain illustrates this point.

EXAMPLE: Ain

Ain: Isolated form (ع) and Initial, Medial, and Final forms (ععع)

Positional variants for all letters which take them can be seen in Table 4 on pages 11 and 12.

Table 4 Positional variants of letter shape depending on position in word

	Position in word		
Independent	**Initial**	**Medial**	**Final**
ا	-	-	ـا
ب	بـ	ـبـ	ـب
ت	تـ	ـتـ	ـت
ث	ثـ	ـثـ	ـث
ج	جـ	ـجـ	ـج
ح	حـ	ـحـ	ـح
خ	خـ	ـخـ	ـخ
د	-	-	ـد
ذ	-	-	ـذ
ر	-	-	ـر
ز	-	-	ـز
س	سـ	ـسـ	ـس
ش	شـ	ـشـ	ـش
ص	صـ	ـصـ	ـص
ض	ضـ	ـضـ	ـض
ط	طـ	ـطـ	ـط
ظ	ظـ	ـظـ	ـظ
ع	عـ	ـعـ	ـع

	Position in word		
Independent	**Initial**	**Medial**	**Final**
غ	غـ	ـغـ	ـغ
ف	فـ	ـفـ	ـف
ق	قـ	ـقـ	ـق
ك	كـ	ـكـ	ـك
ل	لـ	ـلـ	ـل
م	مـ	ـمـ	ـم
ن	نـ	ـنـ	ـن
ه	هـ	ـهـ	ـه
و	-	-	ـو
ي	يـ	ـيـ	ـي

Source: Adapted from Lebanese Arabic Institute (2017).

In one study, 25 children in Grade 6 and 10 adults were asked to name real pronounceable Arabic words written with an incorrect letter variant, which maintained the phonology but changed the orthography of the words (Taouk & Coltheart, 2004). Although children and adults were equally accurate, the adults read the words more slowly. The presence of incorrect positional variants adversely affected their word reading ability to a greater extent. Other important points from the same study are referred to later in the thesis.

2.3 Word structure in Arabic

Word structure is an important feature of colloquial Arabic and MSA. This section deals with word structure in MSA under three subheadings: morphology, verbs, and nouns, all of which provide assistance in reading printed Arabic text when short vowels are not shown. Some understanding of how certain features of word structure are represented in Arabic orthography will help to shed light on what information is available to skilled readers in the absence of short vowels. Verbs and nouns have been chosen to illustrate these features because they are the same kinds of items which are used in the studies and experiments reported in this thesis.

Reading texts rather than isolated words in English is considered to involve several processes related to the orthography, phonology, meaning and grammatical structure, in addition to drawing on experience of previous discourse (Balota, Paul, & Spieler, 1999). Many scholars (e.g., Frost, 2012) argue that processing of morphological information is also important.

The intention here is to not to explore Arabic morphology in depth, but to acknowledge that there are relationships between orthographical, phonological and morphological representations of language which affect the cognitive processes involved in reading (Boudelaa & Marslen-Wilson, 2001; Frost & Katz, 1992). Differences in these relationships have implications for understanding the processes of visual word recognition and word naming in reading because they affect how the lexical representations are stored and accessed (Boudelaa, 2014).

The relationship between orthography, phonemes and morphemes is not the same in all languages. In Mandarin Chinese, for example, the characters represent morphemes and syllables rather than distinct phonemes. The structure of Mandarin Chinese can be termed morphosyllabic (McBride-Chang, 2004). In Spanish, graphemes and phonemes correspond in an alphabetic writing system (Tolchinsky, Levin, Aram, & McBride Chang, 2012), whilst Hebrew and Arabic have a largely consonantal writing system or abjad.

When Venezky (1970) reported his findings regarding English orthography, he identified that words were organised according to what he termed the morphophonemic principle. Charged with designing a spelling curriculum, he examined first the historical development of English spellings from a range of other languages, then investigated morphemes. He found that separating words into their morphemes was necessary before regular correspondences between graphemes and phonemes could be identified. Thus, English spelling represents morphemes as well as phonemes, and the representation of morphemes is consistent although the pronunciation may vary. This implies that words should be separated into their morphological components before consideration of phonological patterns.

2.3.1 Word structure in Arabic: morphology

In English morphology, prefixes and suffixes are added to word stems to produce different grammatical forms such as plurals or the past simple tense, and to derive different word types from a single stem, for example ‘un-like’, ‘like-able’, ‘like-ness’. In Arabic, morphological structures and processes are more complex. Conjugations of verbs are formed by the addition of a prefix or suffix or both, while nouns have different endings according to gender and number and some types of pronouns, prepositions and conjunctions can be attached to nouns.

There is continuing debate among experts about how the overall core structure of Arabic words should be described. The options are: root and pattern (three and sometimes four consonants and a vowel pattern), stem (the part of the word that does not change when inflected such as ‘reduc’ in ‘reduced’ and ‘reduction’), lemma (the base form, for example ‘reduce’ in ‘reduced’), and etymon (two of the three, or four, consonantal roots).

This section justifies the choice of a root and pattern approach to morphology in the present study in the light of the more complex morphologies of Semitic languages, and the results of experiments conducted in both Hebrew and Arabic. Two main contrasting theories are used to explain Arabic morphology, one based on concepts of roots and patterns and the other based on the concept of a stem with affixes (Boudelaa, 2014; Mahfoudhi, 2007). The first of these, which is the more widely adopted of the two, assumes that words are constructed of a root and word pattern (Ratcliffe, 2013). Roots and word patterns are abstract concepts which are typically understood as two morphemes, a consonantal root carrying core semantic information and a word pattern or template which contains the phonological and morpho-syntactic characteristics of the word (see e.g., Boudelaa & Marslen-Wilson, 2001; Ryding, 2005; Saiegh-Haddad & Henkin-Roitfarb, 2014). The word patterns consist of long vowels, short vowels and other letters. Roots and patterns are represented quite consistently by the Arabic abjad.

The second theory assumes that most, if not all, regular morphological structures in Arabic can be effectively described in terms of stems that are acted upon by derivational processes (Benmamoun, 2003; Ratcliffe, 2013). However, research in the field of computational linguistics has shown that, in contrast to the ease with which the stem can be extracted in English, it is difficult to extract stems correctly from Arabic words. The process requires extra steps and complex algorithms which take account of infixes and circumfixes (or transfixes). The algorithms are based on identification of the root and pattern before undergoing transformation which involves generating all alternative vowelised versions of the word and determining the most likely particular alternative in a given context, after which the stem is extracted from the most likely alternatives (e.g. Yaseen & Hmeidi, 2014).

As Boudelaa (2014) explains, different definitions of the stem have been used on the basis that some word patterns for nouns do not carry any grammatical meaning. Certain scholars have chosen to use the root of the imperfective tense of the verb as the stem, for example Benmamoun (2003). It can be argued that both nouns and verbs can be created from this stem and the theory has the added advantage of being compatible with explanations of morphology in English and other Indo-European languages (Boudelaa, 2014).

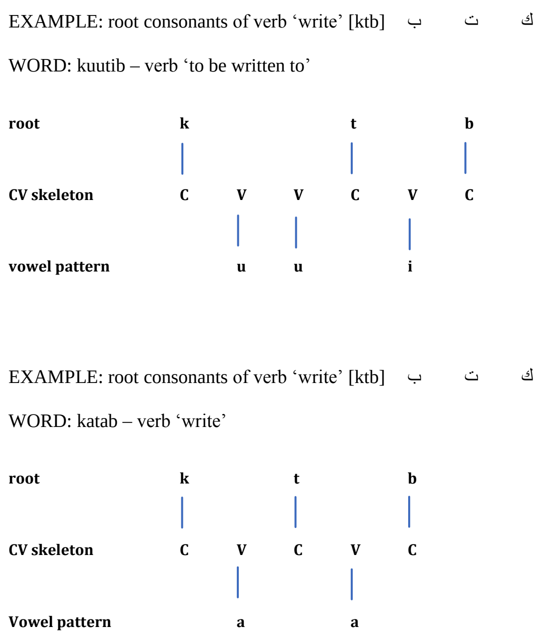
However, morphology in Semitic languages appears distinct. Although Hebrew and Arabic have different orthographies, their morphologies are similar (Ravid, 2012), and Hebrew words are generally considered to be formed by combining a root with a pattern (Ibrahim, 2008). Like Hebrew, Arabic morphology is non-concatenative (Boudelaa & Marslen-Wilson, 2001), productive, and, as discussed in the sections on verbs and nouns, is both derivational and inflectional (Abu-Rabia, 2007). The use of all three morphological processes in derivations can create problems and ambiguities (see e.g., Baluch, 2005), in addition to those for Arabic described in the following paragraphs. Further ambiguities arise from the existence of a colloquial language alongside the formal written language because morphologies of local forms of spoken Arabic have similarities and differences with MSA (Baluch, 2005; Saiegh-Haddad et al., 2011).

A number of studies have presented quite convincing evidence for the existence of morphological structure in the way in which complex word forms are represented and stored in the mental lexicon (e.g., Boudelaa & Marslen-Wilson, 2001; Boudelaa & Marslen-Wilson, 2015; Idrissi & Kehayia, 2004). If the widely accepted view of Arabic as a root-based language is correct, morphological awareness could be expected to make a greater contribution to the cognitive processes of reading in Arabic than to reading in English. English language studies have indicated that morphological awareness is positively related to the development of spelling and reading (e.g., Deacon & Kirby, 2004; Singson, Mohany, & Mann, 2000). Deacon and Kirby (2004) conducted a longitudinal study of more than 100 Canadian children from grades 2 to 5. Results indicated that the influence of morphological awareness was greater on reading comprehension than on reading isolated words, although its influence on reading pseudowords was as strong as its influence on reading comprehension (Deacon & Kirby, 2004). Saiegh-Haddad and Geva (2008) and, more recently, Ahmad, Ibrahim and Share (2014) identified a similar relationship in Arabic, thus this relationship may be important in investigations into word recognition in unvowelised Arabic.

In contrast to the fully vowelised text printed for children up to about the 6th Grade and for older beginners, Arabic texts for adults and advanced readers are published without short vowels (unvowelised). Unvowelised printed text is considered opaque due to the missing vowels, although the presence of morphological information contained in the consonantal roots, long vowels and word patterns has led to proposals that morphemes may play an important role in reading in Arabic (Boudelaa & Marslen-Wilson, 2015; Saiegh-Haddad, 2013). It has been suggested that initial lexical access is via the root which carries the core meaning. Specific semantic information is provided by the phonological pattern together with the root (Abu-Rabia & Abu-Rahmoun, 2012).

This raises questions about how the correct morphemes can be identified from unvowelised printed text in which short vowels are missing from the word pattern morpheme and the root may be interwoven with and accompanied by other clitics (Boudelaa & Marslen-Wilson, 2010). The scale of the challenge is illustrated by the existence of 5,336 roots and 2,324 word patterns in use in Modern Standard Arabic (Boudelaa & Marslen-Wilson, 2010).

Morphological units are not combined in a linear manner, but are formed by interleaving the letters of the two independent morphemes (the root and the whole-word pattern). Thus, the order of root letters is determined by the whole-word pattern and the way in which the root consonants can be inserted (Abu-Rabia, 2007). The distinctions between consonants of the root, whole-word pattern, inserted vowels and other consonants may indicate that additional steps are required for phonological processing. Straightforward examples of how Arabic roots and vowel patterns intertwine in a whole-word pattern are presented in Figure 1.



Notes: CV skeleton = consonant and vowel arrangement on which whole word is constructed.

Figure 1 Intertwined root consonants, CV skeleton and short vowels in word pattern

Figure 1 includes the concept of a ‘skeleton’ consisting of an arrangement of consonants and vowels to which the root and vowel pattern can be ‘attached’ (Boudelaa and Marslen-Wilson, 2004b). Altogether there are 15 verbal word patterns with different combinations of consonants and vowels which are encountered very often in Arabic, each one fixing the inflectional pattern and thus determining word characteristics such as tense and gender (Al-Dahdah, 1989). The verb roots in the verbal word pattern contain semantic information, and can produce new words and meanings based on the particular root. In addition, there are nine word patterns for nouns (nominal word patterns), some more frequently encountered than others. Derivations are constructed from nouns in two ways, by adding either a nominal pattern of the base root or a phonological pattern to the verb tense (Al-Dahdah, 1989).

Affixes to the basic patterns provide further information regarding grammar, but also add to the structural complexity of printed words. The following section describes and illustrates the role of verbal affixes. The importance of word structure in Arabic for the present study lies in its relationship with orthography and phonology as discussed later in section 2.3.4.

2.3.2 Word structure in Arabic: verbs and the role of affixes

Difficulties in defining Arabic morphology as morpheme or stem based are related to the fact that morphology can be derivational or inflectional, and in Arabic it is clearly both (Abu-Rabia, 2007; Ryding, 2005). Derivational morphology refers to the way in which words are formed, while inflectional refers to how words interact with syntax by indicating attributes such as case and tense (Ryding, 2005). The derivational morphology comprises: the short vowels added to roots without breaking the orthographic sequence; vowels inserted between root consonants (infixes) which disrupt the orthographic sequence; and prefixes or suffixes which together with the roots carry particular meaning-related information (Frost, Forster, & Deutsch, 1997). In contrast to these root and pattern arrangements which comprise derivational morphology, inflectional morphology is represented by attaching prefixes and suffixes to words to indicate person, gender, time and number.

Verbs in Arabic are conjugated for tense, mood, person, gender and number. The way in which they are conjugated depends on the class of verb they belong to, which in turn depends on how they are structured phonologically (Ryding, 2005). The different conjugations are identified in the orthography by addition of affixes and suffixes. Arabic verbs have two types of conjugation, one requiring the addition of both prefixes and suffixes, and the other requiring only suffixes (Baayen, 1994). The conjugations depend on whether the tense or mood refers to completed actions (the past or perfect) or to actions that are not yet complete (the imperfect or present). The past tense is conjugated by the addition of suffixes, whereas the imperfect tense is conjugated by the addition of prefixes and, in certain situations, suffixes as well as prefixes (Ryding, 2005).

Verb conjugation is illustrated by examples from the verb 'write'. Examples use either the third person singular or first person dual plural, whichever illustrates the point more clearly. There are three important points to note. Firstly, subject pronouns (such as 'I', 'you', 'they') are included with the tense of the verb as a single word. Secondly, Arabic has two forms of plural for 'we' and 'you'. If two people are the subject ('you' singular and 'I'), the plural form is referred to as the dual form. The term 'plural' is reserved for three or more people as the subjects. The third point is that most verbs are inflected by gender, but the dual plural is not. In order to avoid presenting a bewildering array of affixes that results from these complexities, the examples have been carefully chosen to illustrate the use of a suffix, a prefix, and a combination of one suffix and one prefix. Reading the examples from left to write shows the additions. The root consonants are shown spaced apart at the start of the examples because they are easier to distinguish.

EXAMPLE: root consonants of verb 'write' ب ت ك

(i) Suffix added to create the past tense of *write* for the dual plural *you* form

كتبتُ	كتبتما
I wrote (katabtu)	you wrote dual (katabtuma)

(ii) Prefix added to create the imperfect (present) tense

كَتَبَ	يكتب
he wrote (kataba)	he writes (yaktubu)

(iii) Suffix and prefix added to create the imperfect (present) tense

كتبتما	تكتبان
you wrote dual (katabtuma)	you write dual (taktubaani)

2.3.3 Word structure in Arabic: nouns and the role of clitics

Arabic clitics can pose a challenge in psycholinguistic studies because they are not only orthographic features but also have syntactic and morphological functions (e.g., Attia, 2008). Clitics are separate morphemes that act as a word and are added to denote a direct object or pronoun, but are always attached to a following or preceding word (Boudelaa & Marslen-Wilson, 2010), because they are dependent on the word for pronunciation in a similar way to 'll' in the English 'we'll' (Saiegh-Haddad & Henkin-Roitfarb, 2014). Up to four clitics can be added as a chain to the front of the word (proclitics) and up to three clitics at the end (enclitics). In addition, proclitics are used to indicate words such as ‘and’ and ‘with’ (Boudelaa & Marslen-Wilson, 2010) and some proclitics can be concatenated.

Most nouns also indicate gender and number by different letters, and have case endings according to whether they are the subject or object of a verb (Saiegh-Haddad & Henkin-Roitfarb, 2014). Enclitics are used for the possessive or genitive case, as for example in ‘the tail of the dog’ (Boudelaa & Marslen-Wilson, 2010; Saiegh-Haddad & Henkin-Roitfarb, 2014).

Some prepositions can be added in a similar way, for example in ‘waiting for the bus’, ‘forthebus’ is written as a single word in Arabic. A number of other words such as ‘the’ and ‘and’ can also be added as clitics.

Following examination of some of the main features of Arabic orthography, letter-to-sound correspondence and word structure, consideration is given to how they relate to each other, in order to gain insight into how printed text is read in the absence of short vowels.

2.3.4 Relationship between orthography, phonology and morphology in Arabic

Although fully vowelised Arabic text has a one-to-one letter-to-sound correspondence, the relationship is no longer transparent once the short vowels are removed. In contrast, internal word structure is clearly represented in the consonantal root, word pattern, affixes and clitics. This indicates a transparent relationship between the unvowelised orthographic form and the morphology of a word. The orthography always fully represents the root, which is considered a key component of lexical access (Saiegh-Haddad, 2013). Even when the short vowels are omitted, the consonants and long vowels of the word pattern are always shown (Saiegh-Haddad, 2013). If the root and word pattern morphemes are easily identified from unvowelised printed text, this will facilitate completing the word with missing vowels, which will in turn give access to the complete phonological representation and thus enable pronunciation of the word (Saiegh-Haddad, 2013). From the point of view of the current thesis, this is an over-simplification in that it does not explain how one particular pronunciation will be selected from among several where only the short vowels are different.

Due to the intertwining of roots and vowel patterns, there is no direct correspondence between morphemes and phonetic units, therefore the full pronunciation has to be deduced from the underlying structure and available information (Boudelaa, Pulvermüller, Hauk, Shtyrov, & Marslen-Wilson, 2009). The distinctions between consonants of the root, vowels, and other consonants, may indicate possible distinctions in how they are processed phonologically.

2.3.5 Phonology in English, other languages and Arabic

The correspondence of graphemes to phonemes varies according to the language. Frost and Katz (1992) asserted that in English, the extent to which phonology was directly represented in spelling was somewhere between Hebrew (complex) and Serbo-Croatian (straightforward) and that this reflected a combination of sound-to-letter consistency with morphological relationships which resulted in sound-to-letter inconsistencies. In Chinese, there is no sound-letter correspondence and several characters can be pronounced in the same way; however, there are orthography-to-phonology correspondences at the character and syllable level (Tan & Perfetti, 1998). In Arabic, the one-to-one sound-to-letter relationship in fully vowelised text disappears when short vowels are removed because each consonant can be pronounced in three different ways (as three different phonemes) according to the short vowel which follows it in a particular word, and therefore the full form of a word may have to be recognised before it can be pronounced correctly.

As Blythe (2014) noted, word recognition in English may involve several steps: decoding, precoding and recoding. Decoding is often used to describe the conversions of graphemes to phonemes which are necessary for word identification and is typically associated with young children who are in the very early stages of learning to read. Recoding refers to the fast processing by adults of the phonology of printed words, which occurs in silent reading, and precoding describes the very rapid process of initial consideration of possible forms of the word before eye fixation occurs. In Arabic it has been proposed that readers may have to break down letter strings into their component morphemes in order to identify the word and its pronunciation by inserting the missing short vowels and accessing the phonemes (Boudelaa, 2014).

In view of this, it can be said that reading printed words which look identical due to the absence of short vowels may be more complex and challenging, in terms of the cognitive processes involved, than reading words which have only one possible phonological interpretation. The complexity and challenge may arise from the number of different possible alternative fully vowelised forms of a given letter string. When children learn to read and spell in Arabic, they learn to pronounce a consonant with its accompanying diacritic, so that the word for ‘door’ which is ‘bab’ in Arabic is pronounced as ‘ba-b’ and not as ‘b-a-b’. The word for father is pronounced ‘a-bu’ and not ‘a-b-u’. In the spelling of these two words, the diacritic together with the letter shape tells the reader whether the correct pronunciation is ‘ba’ or ‘bu’. Without the diacritic, the reader has to identify the word in order to know which of the three possible pronunciations determined by the short vowel is correct. Intuitively, the decision may be based on context, or, in the case of isolated letter strings, the most frequently encountered alternative either over their lifetime or as a result of their particular circumstances, or perhaps the first alternative that they learned.

2.3.6 Homographs in Arabic

In Arabic, for many words, different words that have distinct spellings when written with diacritics can appear identical when written without short vowels. The resulting ambiguity has been supposed to necessitate greater reliance on context to disambiguate the lexical entries corresponding to unvowelised letter strings, in reading, to permit visual word recognition prior to establishing meaning through comprehension processes (Abu-Hamour, 2013; Abu-Rabia, 1997a; Saiegh-Haddad, 2004).

EXAMPLE:

wrote (kataba) كتب is identical to *was written* (kutiba) كتب

It can be seen from the vowelised forms that differences are small and require good visual perception skills:

wrote (kataba) كَتَبَ *was written* (kutiba) كُتِبَ

According to Abu-Rabia and Taha (2006), homographs are abundant in standard printed Arabic text, possibly so much so that every second or third word is a heterophonic homograph. Arabic is considered to have many more homographs than Hebrew (Ibrahim, 2009a) and it has been estimated that in Hebrew almost one in four words in unpointed text are homographic if seen out of context (Shimron & Sivan, 1994). An equivalent example in English would be 'bll' which could be 'ball', 'bell', 'bill' or 'bull'. 'He rang' or 'he paid' would make it quite clear which meaning was intended. In Arabic, 'he paid' would be one word and 'the bill' would be another word.

It has been argued that a list of words without diacritics and lacking context should not make sense because, despite looking identical, they could entertain a number of different meanings (Abu-Rabia, 1997a). This is because *"…when short vowels are not present, sentence context becomes a crucial factor for disambiguating homographs…"* (Abu-Rabia & Taha, 2006, p.325). However, as suggested earlier, the correspondence between morphology and orthography makes it easier to identify at least one possible alternative for each letter string. In the absence of a disambiguating sentence context, if people are able to read aloud isolated heterophonic homographs, they must have another way of reaching the specific word form required.

Early psycholinguistics research in Arabic indicated that skilled readers read without the help of short vowels (diacritics) and use grammatical and semantic information to disambiguate heterophonic homographs when they appear in printed text (Abu-Rabia, 1997a; 1997b; 1998). If presented with a single word without diacritics, readers have to work out the diacritics, or rely on their existing knowledge of words, or both. The relatively high proportion of ambiguity in written Arabic may mean that readers have to deduce the word and its full word pattern from the root consonants which are present (Boudelaa, 2014; Boudelaa & Marslen-Wilson, 2005). Whether a word is homographic or not is therefore highly significant because different cognitive processes and routes of access to the mental lexicon may be involved. For this reason, homography will be considered the most important feature and focal point of this study in terms of conducting a name agreement survey and the word naming and lexical decision studies.

The prevalence of homographs may mean that readers have to rely more on other information when decoding unvowelised words. Arabic is considered to be a transparent orthography when written with diacritics and a deep orthography when written without diacritics (Taouk & Coltheart, 2004). Many studies have reported that adult readers have to use their existing vocabulary and previous reading and knowledge of literary Arabic, in addition to morphology (e.g., Abu-Rabia, 1998; Abu-Rabia, Share, & Mansour, 2003) in order to correctly read and understand the written word. They may access a word in the mental lexicon they have seen before, possibly many times, which would fit the context, or they may use their knowledge of the core meaning given by a consonantal root and their knowledge of how words are structured to infer the meaning. For example, knowledge of word patterns, stored separately, could determine whether the word was a noun or verb.

Two types of homograph are found in Arabic, heterophonic and homophonic. Heterophonic homographs occur only in the absence of diacritics. The series of letters which appears on the page or screen affords numerous possible combinations of diacritics (Taouk & Colheart, 2004) which would, in turn, typically yield a different pronunciation for each combination.

For example, the letter string كتب unvowelised/without diacritics allows a number of possibilities including:

كَتَبَ KaTaBa (he wrote)

كُتِبَ KuTiba (it was written)

كَتَّبَ KaTTaBa (he forced (someone) to write)

Homophonic homographs are words with the same spelling, even with diacritics, and pronunciation, but which have different meanings (Taouk & Coltheart, 2004). An example of this can be seen in the word بَيْت *BaYT* which can mean either a house or a couplet in poetry.

The use of root morphemes to construct words leads to frequent occurrence of homographs when short vowels are omitted, and many words will share the same sequence of consonants (Taouk & Coltheart, 2004) even though they can be read in a given context as distinct lexical entities (Azzam, 1993). Share and Levin (1999) have asserted that, in Hebrew, affixation is one of the causes of homography and places additional demands on readers. An unvowelised string can be read in different ways according to the way in which the reader completes the word. Diacritics not only determine the pronunciation of a letter string but carry syntactical information. Diacritics are different for nouns and verbs, and therefore a consonantal string which lacks diacritics can be read as either a noun or a verb. Moreover, many such letter strings can be read as more than one noun and verb, which have different morphological features and belong to different parts of speech.

The database of 1,474 words described in chapter four has been developed with the aim of harnessing some of the morphological resources of the Arabic language. Heterophonic homographs (letter strings with multiple vowelisations, pronunciations and meanings) and non-homographs are included. For homographs, name agreement has been used to determine which vowelisation, pronunciation and meaning can be considered correct (dominant) for use in the experiments in this study and also in future psycholinguistics experiments. The database will permit detailed investigation into disambiguation of homographs in Arabic, thus revealing more about the relationship between morphology, phonology and orthography and the cognitive processes involved.

Chapter 3 LITERATURE REVIEW

3.1 Introduction

Reading requires the coordination of a complex set of perceptual and cognitive processes (Rayner & Reichle, 2010). These cover a spectrum from low-level visual perception to the control of eye movement, phonological processing, detection of word forms, and the higher-level linguistic processes involved in accessing meaning (Norris, 2013). The complexity of these processes is increased by the existence of words with multiple meanings and the associated requirement to select the appropriate meaning (Peleg & Eviatar, 2008). In Semitic languages, the processes for disambiguation of meaning are further complicated by the presence of a complex, productive morphology (e.g., Wintner, 2009), which, in the printed word, is reflected in a homographic orthography (Abu-Rabia & Siegel, 2003; Taouk & Coltheart, 2004). It is argued that readers require at least implicit knowledge of phonology in order to be able to insert the missing short vowels and also knowledge of syntax because the vowels to be inserted typically depend on grammatical features such as gender, tense and number (Abu-Rabia & Abu-Rahmoun, 2012). This may be an overstatement of the knowledge required because very young children learn, without any formal instruction, a spoken variety of Arabic which has its own phonology and similar grammatical and morphological structures. Pronunciation of words may be learned and remembered without association with the written form (Coltheart, 2006). However, in at least one Arabic dialect, Jordanian Arabic, ambiguity results from the inter-relationships between the formal written Modern Standard Arabic and the spoken dialect, a particular feature resulting from diglossia, the situation where a colloquial language variety is used at home and in general conversation alongside a formal variety which is used in specific settings such as education and official communications (Salim, 2013).

The individual processes and the way in which they interrelate to facilitate reading continue to be a matter of scholars' interest and investigation (e.g., de Groot, 2013). Extensive research has been conducted over time into how the characteristics of words affect reading performance in English and other languages (e.g., Coltheart, 2012; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Seidenberg & McClelland, 1989). Much of this research has been based on lexical decision and word naming tasks (Ferrand et al., 2011). The lexical decision task requires participants to quickly decide whether a letter string displayed on a computer screen is a word or not, while the word naming task requires participants to quickly read aloud a word shown on the screen. In comparison with English, relatively little research has been undertaken into the word characteristics that determine Arabic reading performance, although there have been several main strands of interest: the role of vowelisation and other characteristics of the orthography (e,g. Abu-Rabia, 1997b; Ibrahim, 2013), morphology (e.g. Boudelaa & Marslen-Wilson, 2005; Saiegh-Haddad, 2013), the development of psycholinguistics resources (e.g. Boudelaa & Marslen-Wilson, 2010), and diglossia (e.g. Ayari, 1996; Saiegh-Haddad, 2004).

This literature review examines findings, themes and issues from existing research into the impact of word characteristics on reading in English, Arabic and other languages in order to understand the areas of importance for investigating the effect of printed word attributes on Arabic reading. The chapter begins by examining key variables relevant to the present study that affect word identification, reading aloud and lexical access, including word frequency (Balota, Cortese, Sergent-

Marshall, Spieler, & Yap, 2004), word length (Ferrand et al., 2011), neighbourhood density and orthographic depth (Katz & Frost, 1992), name agreement (Snodgrass & Vanderwart, 1980) also termed meaning dominance (Gawlick-Grendell & Woltz, 1994; Gee & Harris, 2010), and AoA (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Morrison & Ellis, 2000). This is followed by consideration of some of the key cognitive processes involved in word recognition and reading aloud, the models that seek to account for them and how the models explain the effects of variables. This is followed by discussion of the purpose and nature of laboratory tasks used in this study, and the chapter concludes with a critical assessment of relevant psycholinguistic studies in English, Arabic and selected other languages.

3.2 Variables

One of the challenges in psycholinguistic research is related to the nature of corpora, which are intended to be representative of a specific linguistic community in one or more situations, such as literature or social media. Corpora contain a complex system of interacting variables which cannot always all be controlled for. A number of variables related to the individual characteristics or mood of the participants in any particular study or experiment also cannot be controlled for. It is therefore important for researchers to control where possible for potentially confounding variables in order to understand how those variables interact (Grondelaers & Speelman, 2007). Rating studies can provide a source of information that can be used to control for one or more variables in a full study (Cortese & Simpson, 2000), and rating study results can be used for analysis (see e.g., Balota et al., 2004). The combined requirements for control, analysis and comparability are reflected in the literature, which indicates that rating studies tend to follow or develop previous models. New items, measures and measurement scales appear to be introduced only when earlier results are shown not to be fit for purpose. The following sections examine some of the key variables in linguistic rating studies, rating scales and statistical analysis of experimental data, bearing in mind that these are largely based on Anglocentric scholarship. Other variables may apply in some languages, for example the inclusion of the number of strokes and components in Chinese characters (see e.g., Liu, Shu & Li, 2007) and vowelisation in Semitic scripts (see e.g., Abu-Rabia, 2001).

Variables that have been the subject of rating studies and experiments include: AoA, name agreement (mainly for pictures), familiarity, imageability, neighbourhood density, subjective frequency and word length. There are a number of frequently uncontrolled variables in psycholinguistic experiments as can be revealed by the particular characteristics of even a single set of 10 to 20 stimuli (Balota et al., 2007). For example, in a lexical decision task in English, many variables have been shown to affect lexical access of words and non-words. The variables concerned include letter and phoneme length, orthographic similarity or regularity, and regularity of similar words in addition to word frequency (see e.g., Cutler & Davis, 2012).

Certain variables have been shown to generally have more influence than others, and there are four in particular which are referred to as ‘the big four’ by Kuperman et al. (2012). The first of these is word frequency, which is especially important in lexical decision tasks, in which it can account for more than 30% of the variance in response times (Balota et al., 2004; Brysbaert & New, 2009). The second is word length, which is measured in characters or syllables and sometimes in both, and has been shown to be influential in word naming in addition to lexical decision tasks (Ferrand et al., 2011). The third refers to the similarity of one word to other words, or neighbourhood density. In one study, stepwise multiple regression analysis revealed that after frequency, the second most influential variable was a measure of neighbourhood density, which accounted for almost 13% of the variance in response times in lexical decision (Brysbaert et al., 2011). The measure used was OLD20 which is calculated as the mean Levenshtein Distance (LD) from a word to the twenty words with the most similar orthography based on the fewest number of additions, deletions, and substitutions of letters (Yarkoni, Balota, & Yap, 2008). In English, this has typically been measured in terms of the average frequency of the letter pairs in the word, or alternatively in terms of the orthographic neighbourhood which has been defined as the number of words that can be formed by changing a single letter in the target stimulus (Balota, Law, & Zevin, 2000), although OLD20 includes comparisons in the lexicon between all pairs of words even if they have different lengths. The appropriateness of measures of neighbourhood density in Semitic languages is examined in section 3.2.3. The fourth important variable, in word naming, is held to be the first phoneme of a word because of its influence on of monosyllabic words, although the first phoneme has little effect in lexical decision tasks (Balota et al., 2004).

In order to give an idea of the relative importance of variables, when these four variables were subjected to stepwise multiple regression analysis, the most influential variable in prediction of lexical decision time was word frequency, accounting for just over 40% of the variance (Brysbaert et al., 2011). Word length measured in syllables accounted for just over 1% in lexical decision, while the combined effect of other variables accounted for 2.0% (Brysbaert et al., 2011). The measured impact of variables is not the same in all languages, or indeed in all experiments depending on their design. The following variables are important for this study, hence the decision to include them as specific items: frequency, word length, name agreement and AoA. Measures of neighbourhood density in the context of Arabic homographs are also discussed.

3.2.1 Frequency

Word frequency refers to how often the word has been encountered. Frequently encountered words elicit more rapid responses than those encountered less often, in both word naming and lexical decision tasks (e.g., Balota & Chumbley, 1984; Balota & Spieler, 1999). Word frequency is typically expressed in the count per million words in a corpus, termed objective frequency, although some studies have used subjective frequency ratings determined by participants on a scale of, for example, 1 to 5 or 1 to 7 (e.g. Frost, Katz, & Bentin, 1987; Vaknin-Nusbaum & Miller, 2011).

Frequency is traditionally understood as meaning objective word frequency, as discussed in articles such as Brysbaert and New (2009), which describe and critique several interpretations of objective frequency whilst acknowledging that measures have evolved over time. They state that the source of the corpus for establishing frequencies is an important factor and that, for a long time, corpora were drawn primarily from books, newspapers and magazines. Despite advances to include television and film subtitles in order to be better representative of everyday language (e.g., New et al., 2007), this type of frequency is clearly objective frequency as it does not involve surveying people to establish the frequency with which they had heard, spoken, seen, and written that particular word. The same is true for most of the Arabic corpora which have been created since 2010 (see e.g., Abdul Razak, 2011). Although objective frequency counts give good estimates of how often particular words appear in printed text, it may also be useful to know how often participants encounter the word, a measure termed subjective frequency. It is proposed by some scholars that this may be more useful than the objective frequency count because most printed word corpora do not take into account spoken word frequency (Cortese & Balota, 2012), although the CELEX database is an exception. (Baayen, Piepenbrock & Gulikers, 1995). Objective frequency may be more appropriate for Modern Standard Arabic, since printed text is where most native Arabic speakers will regularly encounter the words. Colloquial Arabic, which has no written form, is used at home and in the majority of work situations outside formal education. Films and some television programmes are generally presented in one of the colloquial forms.

A previous measure was familiarity, which has proven difficult to define, partly because standard instructions for familiarity ratings have been unclear (Cortese & Balota, 2012) and, in some cases, because familiarity ratings have been shown to be linked to meaningfulness rather than the intended target (Balota, Pilotti, & Cortese, 1999). Hence there have been rating studies looking at a measure of how often participants encountered particular words in speaking, listening, reading and writing. Balota et al. (2004) identified that, in lexical decision and naming tasks, subjective frequency ratings could predict performance over and above objective frequency, neighbourhood size, grapheme-phoneme consistency and word length. It is therefore important to be clear about how frequency is defined and to state how it is measured in order to ensure that results of studies are not attributable to exceptionally low- or high-frequency words.

A particular difficulty in examining the effect of word frequency in Arabic is that there is a lack of systematic data on word form frequency in Arabic (Ganayim, 2015). The best available information to date for the purposes of the present study is obtainable through the Aralex database (Boudelaa, & Marslen-Wilson, 2010), which gives frequency counts for the root and the unvowelised stem, also token frequency, and bigram and trigram frequency. Aralex is a corpus of 40 million words constructed from printed, therefore non-diacritised, Modern Standard Arabic (MSA) text taken from newspapers (Boudelaa, & Marslen-Wilson, 2010). AraMorph (Buckwalter, 2002) was used to produce full morphological analysis, part of speech (POS) tags, and a vowelised version of all possible alternative versions, together with stems, affixes and clitics. The AraMorph outputs were then combined with a set of automated rules covering concatenation as well as disambiguation in order to generate probability scores for the correctness of the version and a measure of the uncertainty attached to that score. Accuracy of the resulting database entries, in terms of whether words were 100% correct according to dictionaries, which resulted from combining the morphological analysis with the automated rules was 85% with case endings included, and 93% when they were excluded. Further cross-checking took place, both with Egyptian experts in Arabic and the Hans Wehr dictionary. Accuracy for the final 40 million word corpus has been given as 90% without case endings, and 80% for complete diacritics. Before Aralex (Boudelaa, & Marslen-Wilson, 2010), frequency was typically controlled for by taking high- versus low-frequency as determined by main and subordinate dictionary definitions of words with only two possibilities, or by expert or participant rating. Since 2010, Khwaileh and Body (2014) have used Aralex to control for orthographic frequency in Tunisian Arabic, which has no written form, while trying to take into account differences in the colloquial form. An extensive search of available literature has not revealed any studies which directly address the effect of frequency on reading printed MSA text in Arabic since then.

The Qur'an Lexicon Project (Binte Faisal, Khattab, & Mckean, 2015) has provided a database of frequency counts and other statistics for 19,286 fully vowelised Classical Arabic words as used in the Qur'an. These statistics could not assist the present study investigating printed MSA text, because the homography disappears when text is fully vowelised and this results in different frequency counts. In addition, the difference between Classical Arabic and MSA is similar to the difference between Chaucer and modern English. The most suitable available measure of frequency for the current studies which use unvowelised letter strings was therefore taken from the frequency of the unvowelised letter string provided in the Aralex database (Boudelaa, & Marslen-Wilson, 2010).

3.2.2 Word length

Word length in Indo-European languages usually refers to the number of letters in a word, but may also be measured in syllables or phonemes. Although word length may account for only a small influence on lexical decision (Brysbaert et al., 2011), there has been debate about its impact on word recognition. For example, effects of word length in English were identified independent of printed frequency, and orthographic neighbourhood size (New, Ferrand, Pallier, & Brysbaert, 2006). In the context of lexical decision tasks, New et al. (2006) analysed effects of word length in a dataset of 33,006 latencies taken from Balota et al. (2002). Lexical decision performance was reported to be facilitated by word length of 3 to 5 letters, unaffected by words of 5–8 letters, but inhibited by a length of 8-13 letters. Inhibitory effects of length for words with 8-13 letters and for non-words have been attributed to the need to compile words from letter features and/or letters (Balota et al., 2006; Coltheart et al., 2001). The speeded word naming study by Balota et al.

(2006) showed that printed word frequency influenced the effects of word length; greater length effects were associated with low-frequency words.

Evidence from Semitic languages is relatively scarce. One study by Koriat (1984) found that unpointed string length affected naming latencies, with five-letter strings taking longer to pronounce than two-letter strings. Lavidor and Whitney (2005) reported a significant effect of word length on response times in a lexical decision task when comparing three-letter and six-letter unpointed words. Weiss, Tami and Tali (2015) found that longer unpointed Hebrew letter strings were read more accurately and rapidly than longer, fully pointed words by skilled Hebrew readers. They also found that typical readers responded more quickly to longer unpointed words than to shorter ones. They asserted this was consistent with earlier research findings that word length effect was greater in transparent orthographies, such as the comparison of reading in English and Welsh by Ellis and Hooper (2001). They suggested that reading pointed Hebrew words required serial phonological assembly which would take longer than reading unpointed text.

Another study found that in Arabic word length influenced eye fixation (Paterson, Almabruk, McGowan, White, & Jordan, 2015). Initial eye fixations on shorter words were central, in comparison with fixation closer to the start of a longer word and lasting longer. In addition, refixation was more likely to occur. Thus, it is highly probable that word length influences reading in a diverse range of alphabetic systems but perhaps in different ways. One factor which may affect the influence of word length is that length as defined by number of letters in Latinate languages is different from length as it appears in Arabic. This is because printed text in Arabic typically uses a proportional font in which there can be considerable variation in the size, shape and width of letters as well as variable between-word spacing. McDonald

(2006) identified that using a proportional font slowed reading times in comparison with a monospace font. Another experiment using Arial (proportional) font and Courier in Finnish found that gaze and fixation times were longer for the proportional font (Hautala, Hyönä, & Arob, 2011). The spatial extent of the word was found to affect the likelihood of skipping words as well as the landing position, thus this may be relevant in Arabic reading.

A further factor in reading Arabic is that visual crowding may result from the cursive text which is associated with a lack of spatial separation and loss of clarity (Jordan, Paterson, & Almabruk, 2010). Another important difference between Latinate scripts and Semitic scripts which may influence the effect of word length is found in readers' eye movements in word recognition of isolated words. It has been suggested that in English, greatest efficiency of word recognition occurs when preferred viewing location (PVL) is close to the optimal viewing position (OVP) which is also between the beginning and middle letters of words (Jordan, Paterson, Kurtev, & Xu, 2010). In contrast, in Hebrew it was found that the PVL was to the right of the word centre but the OVP was at the centre (Deutsch & Rayner, 1999). A similar finding regarding OVP was reported for Arabic (Jordan, Almabruk, McGowan, & Paterson, 2011) while PVL was found to vary from central for 3-letter words, to the right of centre for 7-letter words (Paterson et al., 2015).

Thus, it is important to control for word length, whilst recognising that spatial factors as well as number of letters and eye fixations may interact with these.

3.2.3 Neighbourhood density

Neighbourhood density refers to the number of words which are similar to others in their orthography, morphology or phonology. The most frequently used measure of neighbourhood density is the total number of words that can be generated by replacing one letter or sound in a particular word while leaving the others unchanged (Luce & Pisoni, 1998). Words have typically been defined as orthographical neighbours if they have the same number of letters but one of the letters is different; the measure is referred to as Coltheart's N (Coltheart, Davelaar, Jonasson, & Besner, 1977). The underlying assumption is that visual word recognition in alphabetic orthographies is based on recognition of letter sequences. For low-frequency words in lexical decision tasks, a larger N has been considered to elicit faster responses (e.g., Pollatsek, Perea, & Binder, 1999). It has also been proposed that the relative frequency of higher- and low-frequency neighbours affects performance. Andrews (1989) found by manipulating frequency and Coltheart's N that larger neighbourhoods were associated with more rapid word naming and lexical decision for low frequency words. Further experiments aimed to distinguish the lexical contributions to the effect of neighbourhood size from orthographic influences, first by keeping bigram frequency constant while manipulating word frequency and neighbourhood size and, secondly, by keeping neighbourhood size constant while manipulating the other two variables, in lexical decision and word naming (Andrews, 1992). Andrews therefore proposed that neighbourhood size effects were related to lexical similarity rather than orthographic redundancy and also that the effects were related to lexical access rather than particular processes associated with word naming or lexical decision tasks. It was noted that the lexical similarity required further examination and definition in order to increase understanding of word recognition processes.

A further measure of orthographic similarity was introduced by Yarkoni et al. (2008) to reflect the concept of neighbourhood density. This measure, OLD20 (Orthographic Levenshtein Distance), calculates the least number of letter changes required to transform the target word into 20 others. Thus, a word allocated a value of 1 for OLD20 indicates that 20 other words can be formed by changing, deleting or adding a single character.

A study by Vergara-Martinez and Swaab (2012) used four sets of 33 English words in an event-related potential study to investigate the effect of orthographic neighbourhood density as a function of frequency. High and low frequency and OLD20 measures were compared in a 2 x 2 design with a sample of 22 undergraduates. They found that the frequency effect was greater for words with smaller neighbourhoods, and analysis of brain activity indicated that the effects of orthographic neighbourhood density began and finished sooner than those of frequency. They also observed that any temporary competition among lexical candidates was offset by the domination of a high frequency word, but that in a dense neighbourhood the effect of frequency was smaller (Vergara-Martínez & Swaab, 2012).

Measures based on OLD20 may be less useful in Arabic because they give more weight to substitutions of single letters than to transpositions of pairs. However, pair transpositions are far more common in Semitic languages in vowelised text because the phonemes consisting of consonant plus short vowel move as a single unit (Pollatsek & Treiman, 2015). It has also been argued that measures of neighbourhood density such as Coltheart's N and OLD20 do not make sense in Semitic languages because of the root and pattern word structure. Single letter changes in the root would be *"orthographic neighbours in English lexical space"* but in Hebrew lexical space the words could be far removed from each other (Velan & Frost, 2011, p.142). One example of such a change in unvowelised printed Arabic text can illustrate this. Changing 'l' to 'm' in the roots 'a-d-l' and 'a-d-m' produces a change from the root 'a-d-l' meaning 'to put in order' or 'act justly' to the root 'a-d-m' meaning 'to be non-existent'. In Hebrew, lexical space is said to be arranged by morphological root families rather than by straightforward orthographic structure, so that all words which share a particular root will be grouped together or connected to each other (Velan & Frost, 2011).

Morphological neighbourhood density, or what Boudelaa and Marslen-Wilson (2010) term type frequency or family size of roots, may be a more relevant measure than OLD20 in Semitic languages as a dense neighbourhood may activate many more lexical items. Density in Semitic languages appears to lie in the family size generated from the root and word pattern structure rather than in transposed letters. Consonants, and hence roots, were detected before vowels in visual word recognition experiments by Boudelaa and Marslen-Wilson (2005) who used morphological family size as a control variable in investigating effects of both consonantal roots and word patterns (Boudelaa and Marslen-Wilson (2011). Previous research in Dutch (Schreuder & Baayen, 1997) and Hebrew (Feldman, Pnini, & Frost, 1995) had shown that larger morphological family sizes elicited more rapid responses,

Density in Semitic languages may also be related to the phonological rather than the orthographic similarity of words. Visual word recognition is considered to be influenced by phonological as well as orthographic neighbourhood density (Marian & Blumenfeld, 2006) which may be relevant to word naming of isolated Arabic unvowelised letter strings which are homographs. Velan and Frost (2011) postulated that the way in which words in Semitic languages are structured limits how phonemes and their corresponding letters are arranged due to the phonological patterns. These assist readers to predict which consonant or vowel is most likely to follow the preceding phoneme. Based on a sample of 18 bilingual Hebrew-English university students, 20 Hebrew sentences were presented containing a target word of Hebrew origin, together with 20 Hebrew sentences containing a target word derived from languages outside the Semitic family, and 20 sentences in English containing a target word. Half of the sentences in each group contained the target word with transposition of a single internal consonant and half contained the correct version of the target word. The percentages of correctly recognised words in English was 89% without letter transposition and 82% with transposition. A similar result was found for non-Hebrew derived words, 89% for those without letter transposition and 82% with transposition. In Hebrew, the percentage drop in correct reading was larger, falling from 89% to 68%, indicating that letter transposition had a greater effect on reading morphologically complex words. Velan and Frost (2011) concluded that morphologically simple Hebrew words could be read in a similar way to English, while the Hebrew-origin words could not. Hebrew words appeared to limit the ways in which phonemes corresponded to letters in a way that European languages did not. Phonological neighbourhood density has also been considered as a variable with potential to affect and predict reaction times. In contrast to the facilitatory effect of orthographic neighbourhood density which, in word naming, tends to be limited to low frequency words (Andrews, 1997), phonological density may affect words of varying frequencies.

Unvowelised Arabic letter strings which are heterophonic homographs present a particular problem for identifying neighbourhood density because each form of the homograph may have its own neighbourhood densities in terms of orthography, phonology and morphology. Identification of these, and possibly development of a measure involving all of them, requires a separate and major study.

3.2.4 Name agreement

Name agreement refers to the extent to which different people agree on the name of a specific picture (in picture naming) or the meaning of a particular word. In picture naming, which is where this variable has mainly been used, pictures with higher levels of name agreement are named more quickly than those with lower levels of agreement (e.g. Barry, Morrison, & Ellis, 1997). Similar results have been found in Belgian Dutch (Severens, VanLommel, Ratinckx, & Hartsuiker, 2005), French (e.g Bonin, Chalard, Meot, & Fayol, 2002), Italian (Dell'Acqua, Lotto, & Job, 2000) and Spanish (Cuetos, Ellis, & Alvarez, 1999). In lexical decision in English, words with several related meanings elicited faster responses than those with unrelated meanings, but meaning dominance had no effect (Rodd et al., 2002). Meaning dominance was determined by the word associate most frequently given by participants and was rated from 1 for highest dominance to 0.5 if meanings were given by equal numbers of participants. In word naming in English, experiments comparing ambiguous words with consistent and inconsistent pronunciation and having few or many senses (related meanings) showed that words with related senses elicited faster responses, with a larger effect for inconsistent words (Rodd, 2004). Arabic does not have inconsistent letter-to-sound correspondence; instead, much phonemic information is absent in printed text. Thus, the number of meanings and senses as well as orthographic forms associated with a particular unvowelised letter string is more relevant to the present study. Prior to this study, it was not known whether Arabic readers attached the same number of meanings to a homographic letter string as were given in dictionaries and so there was no ready-made baseline or set of norms to use in the experiments.

In one picture naming study, name agreement was found to predict response times more strongly than word frequency and AoA, as well as concept familiarity, image agreement, imageability, visual complexity, image agreement, number of phonemes and number of syllables (Alario et al., 2004). In the present study involving unvowelised letter strings with a number of different pronunciations and meanings when fully vowelised, it was important to determine norms of meaning frequency as recommended by Gee and Harris (2010).

The extent of name agreement, expressed in terms of alternative names or fully vowelised orthographic forms, is typically measured in one of two main ways, either by percentages or by the H statistic. Percentage measurement uses the highest percentage of participants giving the same meaning and may assign an arbitrary cut-off point such as minimum 65% or 90% of participants agreeing on one particular meaning. The H statistic takes into account all the percentages of participants agreeing on all the possible alternatives they identify, thus measures the spread of agreement; this is explained in more detail in chapter 4, the name agreement study.

3.2.5 Age of acquisition (AoA)

AoA refers to the age at which a word was learned or understood. AoA refers to a particular behavioural event, whereas frequency refers to the number of times the word is encountered or used as an adult or over a lifetime (Zevin & Seidenberg, 2002). AoA is measured in one of two ways, either by asking adults the age at which they first learned or understood the word, which gives subjective norms (e.g., Morrison, Chappell & Ellis, 1997) or by observing the age at which children learn or understand the word for example by naming objects (e.g., Morrison et al., 1997) or by collecting data from children’s books (e.g., Monaghan & Ellis, 2010). Subjective AoA values are more commonly used and are established by asking adults to estimate the age at which they think that they first learned certain words. In some studies, they are asked to use a scale, such as from 1 (learned before 3 years old) to 10 (learned after 16 years old) (Gilhooly & Logie, 1980). In other studies they are asked to give their age in years. Scales and years are both estimates and so cannot be considered as accurate, but estimates in years can give a better approximation than the broader bands.

Words learned at an early age elicit more rapid responses than words which are learned later (e.g., Juhasz, 2005; Monaghan & Ellis, 2010; Kuperman et al., 2012). In English, Cortese and Khanna (2008) found that AoA as a predictor variable was additional to those already identified by Balota et al. (2004). A review by Juhasz (2005) identified that the effect of AoA was found in word naming, lexical decision and picture naming.

High correlation has been found between the effects of AoA and word frequency, because words with high frequency are likely to be acquired very early in life compared with low frequency words. In the case of Arabic, high frequency words which are learned before a child starts school may well have been learned in colloquial Arabic. Words for water, for example, are different in Libyan, Egyptian and Saudi colloquial Arabic, although the word for water in MSA is common to native speakers in all three countries. Although Balota et al. (2004) did not find a significant effect of AoA in their analysis of data from the English Lexicon Project, Cortese and Khanna (2007) subsequently used available subjective frequencies from Balota, Pilotti and Cortese (2001) in their analysis of 2,342 words and found that AoA did indeed predict latencies in word naming and lexical decision tasks after allowing for the other variables identified as predictors by Balota et al. (2004).

The correlation between AoA and frequency was examined by Brysbaert and Ghyselinck (2006) in a review of studies in English, Dutch and French. Findings led them to propose two effects of AoA, one was closely related to frequency effect, while the second effect, seen principally in word association and object naming tasks, was considered independent of frequency and to arise from competition among meanings.

It was considered important in the present study to collect subjective ratings of acquisition ages because MSA is learned in almost all cases only when a child starts school. Mean ages of acquisition might be expected to be higher than in some other languages and AoA might be expected to have a smaller effect.

3.2.6 Interaction of variables

Individual variables not only affect and predict how quickly and easily words are identified and pronounced but also interact with other variables. For example, response times are faster for shorter words, more frequently encountered words and words learned earlier in life. Each variable may make a separate contribution to the result, but the contribution of how, for example, AoA and length interact could be different from the individual contributions of AoA and length. The interaction between two variables indicates that the effect of one variable is different for different levels of the second variable. Thus, the effect of AoA may be given by its individual contribution plus the effect of its interaction with frequency (which could in theory be either an increase or decrease, with earlier AoA associated with either higher or lower frequency effects) plus the effect of its interaction with length (which, again in theory, could have two directions). The interaction indicates whether the effect of AoA will be strengthened or reduced by its interaction with length, depending on whether the interaction and the main effect have the same direction (i.e. both positive or both negative) in which case the overall effect is increased, or whether they have opposite directions (i.e. one negative, the other positive), in which case the overall effect will be reduced. Any analysis of the effect or predictive capacity of variables needs to take this into account (Bauer & Curran, 2005).

3.3 Cognitive processes in word recognition and reading aloud

This section outlines the broad approaches to theories of reading before considering how native Arabic speakers read isolated words lacking diacritics and the implications of observations about MSA reading for accounts of the organisation of the mental lexicon. This is followed by examination of key models of the cognitive processes and how they account for the effects of word frequency, word length, name agreement and AoA. Models which focus on visual word recognition are discussed, followed by models which seek to explain relationships between orthography, phonology and semantics.

Various explanations of how people read have been proposed over time, which can be broadly termed bottom-up, top-down, or interactive (Davoudi & Moghadam, 2015). These have resulted in many models of visual word recognition, reading aloud and silent reading (Besner & Humphreys, 2009), all of which seek to describe the various cognitive processes involved in reading by explaining the ways in which word attributes affect their processing. For example, words which are frequently encountered may be recognised and pronounced more quickly than others (e.g. Balota & Chumbley, 1984), while the pronunciation of English words with irregular spelling which are rarely encountered takes longer (Seidenberg, Waters, Barnes & Tannenhaus, 1984). Words acquired early in life words are responded to more quickly than those acquired later (Morrison & Ellis, 2000). The length of a word, its orthographic similarity to other words and the age at which it was acquired also affect the time taken to pronounce it (Ferrand et al., 2011), as does ambiguity in terms of the number of meanings it possesses, and whether or not the meanings are related (Rodd, Gaskell & Marslen-Wilson, 2000). Word length can facilitate or hinder word recognition and naming, as can orthographic similarity, depending on the way in which these attributes affect each other as well as on the task and the language (Andrews, 1997). The effects of these influences on word naming and lexical decision are discussed together with others in section 3.2.

Bottom-up explanations of how people read aloud are based on serial processing, from left to right in English, which uses grapheme-to-phoneme conversion rules to transform letter strings into their phonological representations (Cortese & Balota, 2014). In reading printed Arabic text, working from right to left, letter to sound conversion requires the additional step of completing graphemes by inserting the appropriate short vowels. Since up to 30% of printed words may be homographic in the absence of short vowels, this step may require additional morphological rules for converting letter strings to a series of graphemes in order to reach a phonological representation. Equally, it may require readers to call on their knowledge of vocabulary to identify the word, although this cannot be accounted for by a strictly bottom-up process. Top-down explanations involve the use of existing vocabulary and background knowledge such as a general idea of the topic to guess the meaning. Whilst bottom-up explanations can be more readily associated with how children learn to read and top-down explanations may provide a more intuitive account of what skilled readers do, neither can explain why frequently countered and early-acquired words elicit a more rapid response than those rarely encountered and acquired later. Interactive explanations propose that reading involves both bottom-up and top-down processing. These explanations assume that information from print and information from background knowledge are processed in parallel.

3.3.1 How native speakers read in Arabic

Coltheart (2006) began his review of two of the main models of reading by setting out how people read English, asserting that understanding how people recognise whole words contributes to understanding how whole sentences are understood in print. People remember the pronunciation of unfamiliar words in a foreign language which they may later associate with a particular letter string. In deciding whether a letter string is a real word or not, people do not search through all the thousands of words in their vocabulary until they find the one which matches a visual representation. He proposed that the mental lexicon is organised into separate systems for orthography, phonology and meaning. Moreover, he suggested that as children learn to read and then become proficient readers, knowledge of letter-to-sound relationships increases, and the system expands. This may not be the complete picture for native Arabic speakers.

As children, they first learn one of the colloquial varieties of Arabic, which have root-and-pattern morphologies although they differ in some details (Boudelaa & Marslen-Wilson, 2013). When children start school, they learn Modern Standard Arabic (MSA), learning pronunciation first, mainly by listening and repetition, and the names of the letters. They next learn that each letter can be pronounced in one of three ways (phonemes); this introduces diacritics because there are now introduced graphemes consisting of a consonant plus a diacritic which correspond to phonemes. As they gain larger MSA vocabularies, and the use of diacritics in printed text is reduced, children develop greater awareness of roots and patterns in that language. Awareness of roots develops before awareness of patterns (Taha & Saiegh-Haddad, 2017). After the development of full morphological awareness, the only access to phonemes from an unknown isolated unvowelised letter string in printed text is via the root and probable word patterns, based on an understanding of how the root and patterns fit together.

Verhoeven and Perfetti (2011) have proposed a similar explanation, which states that a complex orthographic word form is processed by morphological decomposition which involves access to morpho-phonological units. The early learning of phonemes in vowelised Arabic before the more formal development of roots and patterns may suggest that there are separate lexicons for phonemes and for morphological units, the roots carrying information about the meaning and the patterns conveying information about the syntax. In order for roots and patterns to be reassembled to allow access to the phonemes, it has been asserted that there is a role for the CV skeleton which defines how the root and vowelisation patterns intertwine in the whole-word pattern (Boudelaa and Marslen-Wilson, 2004). Boudelaa and Marslen-Wilson (2015) stressed the importance of morphological units and morphological processing in Arabic and raised questions about how the mental lexicon might be structured but did not provide an answer.

Deciding whether or not an unvowelised letter string is a word is a different matter. Knowledge of roots, vowelisation patterns, the CV skeletons and word patterns would be needed to check that the arrangement is valid, but access to phonemes would not be needed. If the letter string was homographic, it might not be necessary to access more than the root, together with any affixes and suffixes, to ensure they were valid, because a number of different vowelisations and pronunciations would be acceptable as words. It is possible therefore that separate lexicons exist for roots, vowelisation and whole-word patterns and CV skeletons as well as the orthographic, phonological and semantic lexicons.

3.3.2 Theories and models of visual word recognition

Over time, various theories and models have been developed to offer fuller explanation of the processes. This section considers theories of word recognition which focus primarily on the visual aspects, before examining models which seek to incorporate relationships between orthography and phonology, and identify the roles of context, meaning and morphology.

Early theories of reading focused on the relationships between features of words and word recognition, and on how a recognised word was accessed in memory or the mental lexicon (Lupker, 2005: Balota et al., 2006). It was not only assumed that such a mental dictionary existed, but also that recognition of a word required recognition of letter features followed by recognition of complete letters and then the whole word. Subsequently it was shown that whole words were recognized faster than individual letters, the ‘word superiority effect’ (Reicher, 1969).

This led to the proposal that letter recognition was influenced by the complete word as well as letter features, as seen in the interactive activation model developed by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982). Interactive activation assumed three levels and two kinds of connection. The three levels are: letter features, letters, and words. One kind of connections facilitates links between the levels, while the other inhibits connections within the same level. When a word is presented, representations which match features, letters, and words will be activated. It is assumed that features either activate or inhibit recognition of particular letters, which in turn either activate or inhibit word recognition (Warren, 2012). Feedback from activated word-level nodes controls the recognition outcome. Over time, repeated use of activation pathways from words to letters and letters to words will strengthen representations at letter and word levels respectively. Lower levels of activation of some words and letters will make some representations weaker than others. IA models have been developed to include how words are chosen and ordered in sentences, as well as grammatical encoding (Dell et al., 1999). Activation of whole words takes longer for words with many letters because a higher number of visual features is involved (Hofmann & Jacobs, 2014). IA accounts for the frequency effect by assuming that more frequently encountered words have a higher resting level of activation and so attain a critical level of activation for recognition more quickly (Hofmann & Jacobs, 2014). Age of acquisition and resolution of homographs could also be explained in terms of higher resting levels of activation. However, this single route to reading words implies a one-to-one correspondence between grapheme and phoneme that cannot account for how irregular words are read and pronounced in a language like English.

The ability of IA models to explain morphological decomposition by segmenting whole words correctly in English was questioned by Andrews and Davis (1999) because it required the recognition of shorter words included in a longer one and assumed that letters were position-specific. It was similarly questioned in Hebrew, in terms of its ability to distinguish between roots with the same letters in a different order in such a way that a particular sequence of root letters would activate that root (Velan & Frost, 2011). Velan and Frost indicated that the model would require root morpheme units to be activated by letter units according to strict limits on relative letter-position, and so an additional layer would need to be added to the model which would allow activation between letters and words to be regulated by morphemic units. Thus, it appears that IA models cannot easily account for how skilled Arabic readers deal with the intertwined morphological units in unvowelised text and homographs.

Later models which focus on visual elements of word recognition and which are potentially relevant to disambiguation of homographs include three which address how letter order affects word recognition: the open bigram model (Grainger & van Heuven, 2003); the spatial coding model (Davis, 2010); and the letters in time and retinotopic space account (Adelman, 2011). The key elements relevant to homographs are outlined in the following paragraphs.

The open bigram model was developed in response to questions about the importance of relative word positions, rather than specific positions, in English to explain how anagrams were resolved (Grainger & van Heuven, 2003). Open bigrams are pairs of letters in order but not necessarily adjacent. For example MAKE = {MA, MK, ME, AK, AE, KE}. This model, like other early developments of existing models, assumed left to right reading with associated eye movements and eye fixation points. Converting the bigrams to a right to left order would solve one issue for Arabic, but important questions remain for word recognition of Arabic homographs. Open bigrams in unvowelised printed Arabic letter strings would generate the root consonants in order and any long vowels, but would not contain the short vowels, which could only be obtained from the CV skeleton and short vowel patterns. Thus, vowelised words of different lengths could be produced from the same letter string, depending on the number of short vowels omitted which in turn depends on grammatical inflections as well as the CV skeleton. Both imply a greater dependence on lexical knowledge and access to the lexicon.

The spatial coding model attempts to overcome these difficulties by proposing that a model of letters without any specification of context or position could be set inside a model in which lexical selection occurred on the basis of competition. Input of a word containing the letter ‘f’ would activate the node for that letter irrespective of its position in the word. For example, ‘f’ would be activated by ‘for’, ‘soft’ and ‘loaf’ (Davis, 2010). The position of the letters reflects the space they occupy rather than indicating importance and is uncertain in the early stages of decoding. So Arabic root consonants would each be assigned a certainty of 1 as they were progressively recognised, while certainty about the position of other letters would be less until matched in the lexicon with a known word. Davis further argues that phonological and semantic feedback can be incorporated into the equations that underpin spatial coding. Word frequency is taken into account in lexical decision, although neighbourhood density has no effect according to this model.

LTRS (letters in time and retinotopic space), like spatial coding, assumes orthographic processing does not involve position-specific coding channels, but also proposes that a range of bigrams (open, adjacent and non-adjacent) are additionally activated as representational units (Adelman, 2011). This model would be consistent with a stronger influence of roots than spatial coding alone in visual word recognition in Arabic because the first two root consonants could be detected and bigram and single letter coding would be mutually reinforcing.

None of the models examined in this section gave sufficient explanation for the processes involved in reading aloud, the links between orthography and phonology, or relationships with meaning and context, and so different approaches to theory were developed. Rumelhart and McClelland (1982) had included a phonological component in their IA model to account for processing of spoken words and assumed connections between letters and phonemes but acknowledged that further development of the model was required.

3.3.3 Theories and models of word recognition involving phonology and semantics

The main models considered in this section are the Dual Route Cascaded model (Coltheart et al., 2001), and the connectionist models of Harm and Seidenberg (2004) and Plaut et al. (1996). Consideration is also given to the grain size theory (Grainger & Ziegler, 2011). These theories are based on fundamentally different assumptions about how words are stored and processed, and their ability to explain various aspects of reading in Arabic is discussed.

3.3.3.1 Dual Route Cascaded (DRC) model

The DRC model shown in Figure 2 helps to explain reading aloud by its inclusion of phonological output (Coltheart et al., 2001). As depicted in Figure 2, there are two main routes: the lexical route shown on the left, which has two branches (semantic and non-semantic), and the grapheme-phoneme correspondence (GPC) route on the right.

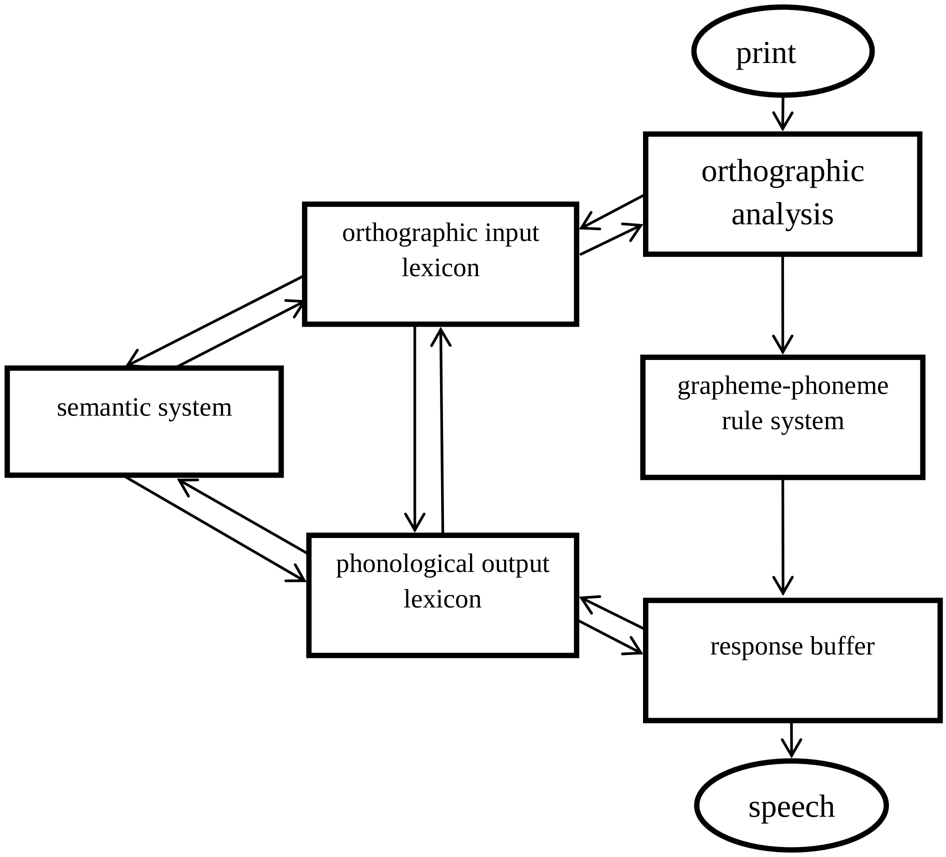


Figure 2 Outline of dual route cascaded model of word recognition and reading aloud
Source: Adopted from Coltheart, Rastle, Perry, Langdon, & Ziegler (2001, 213, figure 6).

Each of the layers depicted in Figure 2, such as the semantic system, contains sets of the smallest individual units of the model, for example, letters and graphemes in the orthographic analysis layer, phonemes in the phonological system, and morphemes in the semantic system. Units in a single layer can interact with each other when the activation of one unit blocks the activation of others. Units in different layers can interact by way of both excitation or by inhibition. Excitation occurs when a signal moves the receiving unit (e.g. letter or phoneme in the appropriate system) towards the threshold at which the unit will become activated. As this happens, the stronger excitation of one unit will enable it to inhibit other units on the same layer of the receiving unit. When sufficient signals are received to reach threshold for action or no action, the receiving unit will respond. Thus, the activation of one unit can influence the activation of others (excitation) or can limit their activation (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

The lexical route assumes that words can be looked up in a mental lexicon which holds knowledge of the spellings and pronunciations of real-word letter strings. The non-lexical route translates graphemes into phonemes using a set of rules governing spelling to sound correspondences and is considered to involve serial activation of phonemes, with each phoneme being activated after the one before. This would be possible in Arabic but could involve many false starts and reruns for unfamiliar words because in many cases the phonemes could only be accessed by identification of the root letters (to access the core meaning) before their associated word pattern, which would reveal the vowelisation and phonemes involved. Ryding (2005) suggests there are 5,000 to 6,000 roots in use in MSA.

The computational DRC model (Coltheart et al., 2001) explains the effect of frequency by assuming that frequently encountered words have a higher resting level than low frequency words, which means that activation is reached more quickly. Any effect of the length of words is minimal and can be explained by the cascade process which, although there is an overlap between one letter being activated and the next, means that words with more letters take longer to process (Coltheart, 2005). AoA is not specifically addressed by the DRC model, and any effect of AoA would probably appear as frequency of encounter with the word over a lifetime. However, this ignores the frequency-independent aspect of AoA identified by Brysbaert and Ghyselinck (2006), and therefore the model would require expansion, for example by a direct link from orthographic analysis to input to the semantic system.

Later computational models have attempted to introduce learning, such as the Learning-DRC (L-DRC) model developed by Pritchard (2013), which incorporates the learning of grapheme-to-phoneme correspondences. In the L-DRC, the resolution of ambiguity in the form of homographs is achieved through competition between the different forms, with activation of the different forms proceeding at different rates so that a 'winner' emerges. However, attempts to train the L-DRC model to learn heterophonic homographs found that some input of context was needed for two meanings in order to achieve activation of a second meaning (Pritchard, 2013).

An alternative approach proposes that the semantic system may be developed by adding a morpho-semantic route (Grainger & Ziegler, 2011). This links morphology to meaning at the highest level with the orthography of complete words on the next level, and the possibility of either single letter or large units such as morphemes. This could help to explain how unvowelised letter strings are read if the model were developed further to allow the whole-word level to contain both vowelised and unvowelised letter strings and the lowest level to contain morphemes as orthographic units.

Name agreement within a DRC model is most easily explained by frequency. High-frequency meanings of unvowelised letter strings are more likely to be encountered by more individuals; lower-frequency meanings will be encountered by fewer people. High-frequency words have a higher resting threshold, and so will be activated more quickly.

3.3.3.2 Connectionist (parallel distributed processing) models

Connectionist models of reading propose that connections exist and operate as neuron-like processing structures. Broadly speaking, there are number of connectionist approaches (e.g. Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996; Harm & Seidenberg, 2004; Monaghan & Ellis, 2010) which share a central concept that information is represented in the variation in weights on connections in a network of simple processing units (Plaut et al., 1996). Interactions between the units are competitive or cooperative and as units respond to input, for example the input of a new word, the weights on the connections vary; this is how the system learns. Information about the orthographic, phonological, and semantic characteristics of words is contained in patterns of activity, and changes in the weights on connections also assume the existence of hidden units that play a mediating role between the patterns. Together, the weights of connections encode knowledge about the inter-item relationships. The collective behaviour of simple processing units in a sizeable network allows complex information to be processed (Seidenberg, 2007). Due to their use in computerised language modelling, they are also referred to in the literature as artificial neural network (ANN) or parallel distributed processing (PDP) models (e.g., Thomas, & McClelland, 2008).

Connectionist models are best considered as networks, with the network understood as a learning device. Within the network, spellings (orthography) and pronunciations (phonology) are contained in discrete groups. The orthographic representations might consist of letters of the alphabet or abjad, or they could be smaller units such as strokes; dots and circles from which letters are formed. In a similar way, the phonological representations could consist of phonemes (for example, the /g/ in peg) or smaller phonetic features based on how the word is pronounced (such as using the mouth e.g., fricative, labial). Strokes, dots and curves can also be termed feature units (Balota, Yap, & Cortese, 2006). Morphological patterns such as the Arabic consonantal root exist as weights on the connections created over time rather than as nodes on the network.

One example of a connectionist model, chosen here because it illustrates some of the key features of such models but is relatively straightforward, is shown in Figure 3 (Harm, & Seidenberg, 2004). The main distinguishing features are an emphasis on processes and processing units; processing units may be categorised as input, output or hidden units. The connectivity refers to the pattern of strengths connecting the individual units and in turn affecting their relative activation states. Another distinguishing feature is the inclusion of the role of context.

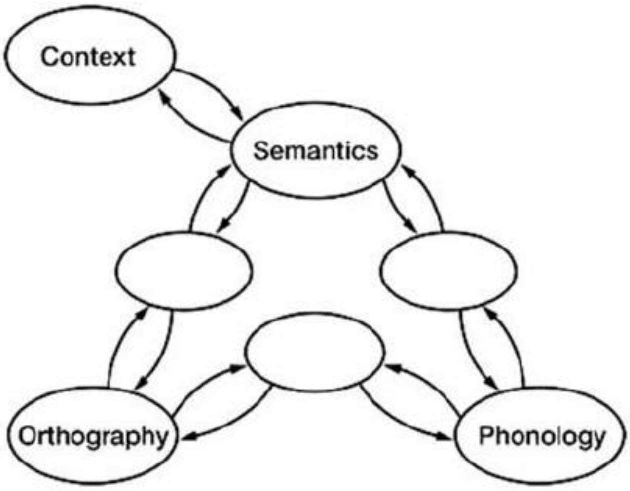


Figure 3 ‘Triangle’ model (Source: Harm, & Seidenberg, 2004, p.663)

Connectionist models typically assume the existence of a layer of ‘hidden’ units (the empty ellipses in Figure 3) between levels of the network which permit the network to learn and hence to be able to represent more complex mapping of words. When a letter string is input to the system in, for example, a word naming task, all the information it holds about correspondences between spellings and sounds is used and this applies whether the word is regular, irregular, ambiguous or an exception word (Plaut et al., 1996). Interactions between the units based on the relationship between the letter string and all words already known, including how they are pronounced, determines the output from the system. Units in the system which match the input are activated and pass through to the output units. For instance, the spelling of the word 'medicine' would be the input, after which a number of units would be activated, and the connections between them, such as phoneme activation, would result in the pronunciation of the word 'medicine' which would be the output.

The flow of activation is controlled by the weights carried by each connection in the network which allows activation to flow in one direction. A weight can be defined as the strength of the connection between units, which is determined by the number of times the units act together, so can be negative (inhibiting) as well as positive (activating). More complex networks are produced by the creation of additional connections between the units on a single layer, new connections between the layers of input and output, and the additional hidden layers containing units that hold information about the context in which the word occurs such as collocations.

Thus, according to the connectionist model, for a particular spelling pattern, the model learns to produce a correct pronunciation of this spelling. This is achieved by the appropriate set of weights being found, which is dependent on several learning principles, some more closely related to how learning occurs in the brain, and other reflecting how learning occurs in AI at a computational level, for instance algorithms.

Examples of learning principles described by Rumelhart, Hinton and Williams (1986) are that learning specifics requires separation from overlapping events and a rapid pace of learning, whereas generalisation requires representations which overlap and a slower rate of learning. The network also need an algorithm which provides instructions about how learning occurs. The most commonly used mechanism for implementing learning is backpropagation. Input data and target output data are provided to the network, and the algorithm calculates the difference between them (the error). This information is fed back to the hidden units and the difference between the input data and target output data is reduced and the weights on connections are altered. The weights on the network are adjusted and the new weights are used for how the network responds when more information is input. As the network learns, the difference between input and target output data is reduced and adjustments become smaller and fewer.

Models like this have been shown to pronounce thousands of words accurately (Seidenberg, 2005). They can account for both words that are governed by rules and the words that are exceptions. Connectionist models assume that, when the input changes as a new spelling pattern is presented, the network employs all of the existing weights to produce activation of units on all layers. The weights are adjusted following each exposure to a word, with larger adjustments being made to connections which contribute more to inaccurate performance. This is done by means of a learning algorithm, a computer program which runs every time after new information is input and matches the weights on the network connections more closely to the input it has received. Performance on a given word is based on one's knowledge of other words, for example the learning of words such as 'love' and 'gate' will result in weight adjustments that will help to pronounce the word 'late'. According to connectionist theories, framed in terms of parallel distributed processing (PDP) (McClelland, Rumelhart, et al., 1986), understanding and fully using the orthographic-phonological correspondences comprise a statistical learning problem which can be resolved through an appropriate algorithm.

The development of connectionist models continues with attempts to address specific aspects of the cognitive processes involved. The Harm and Seidenberg

(2004) model, illustrated in Figure 3, was primarily aimed at explaining how both phonology and visual recognition can contribute to word recognition, including the processing of homophones. There is no requirement for access to lexical representations of a word, only connections via different pathways (the visual and phonological pathways). The model was trained using 6,103 monosyllabic words which included 1047 homophones which represented dominant, subordinate and balanced items in terms of their frequency. Input of sufficient contextual information was needed to retain a subordinate homophone in the network.

A variant has examined how an AoA effect can be reproduced in a connectionist model (Monaghan & Ellis, 2010). The study involved three models: a cumulative frequency, a random frequency and a developmental model. Training the developmental model involved imitating the process of learning to read, from a few words to a vocabulary of 6,229 words. Results of tests using the completed model showed that accuracy rates of reading words which had been entered into the model in the early stages were higher than for words entered later.

In all connectionist models, the effect of frequency is explained by repeated input of a word resulting in greater weights of the relevant connections within the network. AoA can also be explained in terms of earlier encounters with particular words creating the first connections established in the networks as weights to which others are added later. These early inputs create a greater change in weights than subsequent inputs (Zevin & Seidenberg, 2002; Monaghan & Ellis, 2010)).

The effect of length is accounted for by the number of individual connections which are needed; although all the processes operate in parallel, it still takes longer for longer words to be processed. Homographs are more problematic, since they require sufficient input to output mappings of words with inconsistent spelling-to-sound correspondences (Monaghan & Ellis, 2010). This in turn requires sufficient input about the context to generate weights that allow recognition. When there is sufficient input for several possible pronunciations of a single letter string to be recognised as possibilities for output, there is competition among the weights on connections until the sum of all weights generates the output with the greatest weight.

3.3.3.3 Grain size theory

The grain size theory extends the dual route approach by positing the existence of two types of prelexical orthographic codes, one fine-grained code and the other coarse-grained. DRC and connectionist models have been criticised for failing to take into account the extent to which orthography and phonology correspond consistently in differing languages, in terms of their statistical properties and how they are learned (Ziegler & Goswani, 2005; Frost, 2012). This is reflected in differences in reading processes as a result of variations in lexical representation and the development of reading, which led to the concept of a psycholinguistic grain size theory of reading (Ziegler & Goswani, 2005; Frost, 2012).

Grain size theory proposes that correspondences between sounds and written symbols occur in smaller or larger grain sizes, such as letters, phonemes and whole words (Frost, 2012). The fine-grained route is concerned with letter order and start and finish letters, and leads to frequently occurring graphemes being used for converting print to sound and also to prefix and suffix letter combinations being used for morpho-orthographic processing (Grainger & Ziegler, 2011). The coarse-grained route assumes that word identity is derived from minimal letter subsets which give rapid access to semantics. There is some support for this theory from psycholinguistic experiments in Hebrew (Frost, 2006). Citing findings from his earlier experiments, Frost found that in word naming, phonology was worked out from unpointed text, using a fine-grained approach, rather than retrieved as whole words from the lexicon, although lexical access was involved when words were phonologically ambiguous (Frost, 1995). Insertion of missing vowels requires lexical knowledge and the process in Hebrew involves identification of morphemes which control organization of and access to the lexicon (Deutsch, Frost and Forster, 1998).

3.3.3.4 Further developments in theory

Further development of models is ongoing in the search to incorporate individual differences in terms of how and why individuals differ by including sensitivity to factors which affect individual responses (Adelman, Sabatos-DeVito, Marquis, & Estes, 2014). Development of models is also ongoing to fully understand and evaluate the process of reading in languages with very different orthographies. One general shortcoming of lexical access models has resided in their failure to adequately explain how meaning is activated (Balota et al., 2006; Gaskell & Marslen-Wilson, 2002). A greater role for phonology and semantics in recoding strategies has been provided by grain size theory (Ziegler & Goswami, 2005; Grainger & Ziegler, 2011), while attention has also been directed to how readers make correct decisions based on lexical knowledge even when letters are wrongly ordered or missing (Norris & Kinoshita, 2012).

The idea of decisions based on lexical knowledge has been modelled by, among others, Norris (2006) and Norris and Kinoshita (2008) who argue for the role of Bayesian principles of decision making in visual word recognition. These principles use statistical inference to reassess the probability for a particular decision hypothesis as more information becomes available. In word naming, for example, the reader makes a ‘best guess’ based on perceptual evidence such as the root and suffix together with the likelihood of the word (such as objective or subjective frequency). This implies that Arabic homographs are named according to some combination of the CV skeleton, morphological information and frequency.

According to Frost (2012), theories of orthographic processing should therefore form part of a general theory of meaning recovery which also takes into account the interaction between the reader and his or her linguistic environment. The similarities between Hebrew and Arabic morphology (roots and vowel patterns) and unvowelised printed text suggest that Frost's proposals for a system which learns the orthographic, phonological and morphological constraints of MSA, as children do, are appropriate. Homographs would require some contextual input in order to be able to distinguish between multiple meanings, but a solution has not been suggested to meet the challenge of reproducing different pronunciations and meanings for heterophonic homographs, some of which are also homophonic. It is intended that the database of Arabic words which includes heterophonic homographs, and the list of non-words, can be used in a series of experiments which will shed light on this type of interaction.

A different approach has been taken by the search for a model to explain picture naming (Alario et al., 2004). Picture naming shares similarities with word naming of unvowelised letter strings, and so the picture naming model is described here in relation to word naming. The process begins with visual recognition which, when applied to naming unvowelised letter strings, would involve the complexity of the stimulus (number of possible vowelisations) and name agreement according to participants. The individual's semantic system would then process the input. In picture naming this would done on the basis of imageability which, in word naming, could be the same. The next step involves the lexical system where the naming process is affected by frequency, AoA and possibly other word attributes before the word is pronounced.

3.3.4 Comparison of dual-route and connectionist theories

On the one hand, dual-mechanism theories emphasise the differences between the rule and memory subsystems, asserting that these differences extend to governing principles, mechanisms of acquisitions, and types of words; moreover, the systems are located in different parts of the brain. However, dual route theories cannot explain the overlap between rule-governed regular forms of words and exceptions which contain elements of regular forms, such as the example given by Seidenberg (2005) of the exception word 'pint' which shares structure with *'pant'* and *'pine'*.

The DRC model is in agreement with the Orthographic Depth Hypothesis (Katz & Frost, 1992) and the corresponding concept of orthographic transparency (Share. 2008; Saiegh-Haddad & Geva, 2008), which propose that in more transparent orthographies, readers make more use of phonological processes. In less transparent, or deep, orthographies, it may be easier to carry out lexical search using direct association between the visual representation and the word. Both the Orthographic Depth Hypothesis and the DRC model in their traditional forms do not specifically allow for or explain the involvement of morphological factors which, as indicated in Chapter 2, are important in Arabic.

Connectionist theories are underpinned by the assumption that just one set of connections from spelling to sound can explain how people can pronounce both regular and exception words and why reaction times differ (e.g., Plaut et al., 1996). Connectionist models not only imply that phonological access is essential for reading in both transparent and opaque orthographies, but also assume that other factors are involved such as characteristics of the language, and individual reading proficiency (Frost, 2006; Ziegler & Goswami, 2005). Morphology is not directly represented but in a developmental model will appear as patterns formed over time by the weights of connections.

3.4 Disambiguation of homographs

The study of homographs provides an opportunity to examine how words with multiple meanings, depending on their fully diacritised forms, are read and understood. Three main theories have been proposed to explain the role of context in disambiguation where context is available: ordered access (Simpson & Burgess, 1985), selective access (Martin, Vu, Kellas, & Metcalf, 1999), and exhaustive access (Onifer & Swinney, 1981).

According to the ordered access model (Simpson & Burgess, 1985), isolated ambiguous words may be processed in two-stages; initial activation of more than one meaning is resolved by the advantage of the higher frequency word (dominant meaning) and inhibition of subordinate meanings. The rate of activation of a specific meaning depended on its frequency relative to other meanings. Simpson and Burgess (1985) found that in the absence of context the dominant meaning was accessed before the subordinate meaning, and that participants could not control this response even when instructed to identify the subordinate meaning. In the same study, it was found that facilitation for the subordinate meaning declined more quickly than the facilitation for the dominant meaning. This was attributed to inhibition in the form of language processing resources being directed towards a particular meaning and hence being diverted from alternative meanings.

The selective access theory applies to disambiguation of words within sentences; it assumes a context within which meaning is activated and that the meaning accessed depends on the type and strength of the context, together with the word frequency (Martin et al., 1999). The exhaustive access theory proposes that all meanings of the ambiguous item are activated, at least briefly, and such access is independent of context (Onifer & Swinney, 1981). This assumes that access to all meanings takes place before the contextual constraints of the sentence determine the decision. For example, ‘roll’ can only mean bread roll in ‘I ate a roll for breakfast’, whereas ‘roll’ can only mean the action in ‘I am teaching my daughter to roll a ball to me’.

There appear to be differences in how different languages are processed in different languages, as assessed by neuropsychological research. According to Peleg and Eviatar (2008), many studies of lexical ambiguity have investigated semantic organization in the two hemispheres of the brain. In English and other Latin orthographies, homographs are mainly homophonic and the way in which elements of phonology and semantics interact may differ from the interaction of phonology and semantics in heterophonic homographs in a Semitic language. Such differences may be observable in the results of neuropsychological studies (Peleg & Eviatar, 2008). They presented Hebrew homophonic and heterophonic homographs to one or the other hemisphere in order to investigate such differences and compared their findings with a connectionist network model. They reported earlier research such as Grindrod and Baum (2003) which demonstrated that disambiguation necessitated both hemispheres to function correctly. The visual orthography produces more activity in the right hemisphere, whilst the phonological dimension has a greater effect on the left hemisphere. The first (dominant) meaning of heterophonic homographs was the only one to be activated. Disambiguation of heterophonic homographs entailed identifying two or more phonological codes which mapped onto a single set of orthographic symbols. In contrast, homophonic homographs have a single phonological code but more than one meaning, and disambiguation involved activation of subordinate as well as dominant meanings (Peleg & Eviatar, 2008). Differences were observed between Arabic speakers and the Hebrew and English speakers in their responses to a lateralized LD task (Ibrahim & Eviatar, 2012). Earlier experiments involved Hebrew, English and Arabic native speakers in lateralized LD tasks which involved manipulation of the morphological structure of words and non-words (Eviatar & Ibrahim, 2007). These have shown that for Arabic speakers as well as Hebrew speakers, both hemispheres were sensitive to morphological structure. It should not therefore be assumed that all aspects of all languages will be processed in exactly the same way.

In the absence of a disambiguating context in Arabic, homograph processing will involve processes other than reliance on context. Investigating lexical decision for ambiguous consonant strings in Hebrew, Bentin and Frost (1987) found that decisions were made more quickly for ambiguous strings than for vowelised alternatives for the same strings. This applied to both low- and high-frequency vowelised versions, suggesting that these lexical decisions were based on the ambiguous orthography. The word frequency effect was significant for ambiguous as well as unambiguous words. In a word naming task, participants were significantly faster in naming high-frequency words from ambiguous strings than low-frequency vowelised alternatives. Naming latencies were not affected by vowelisation. These results led them to hypothesize that in Hebrew, lexical decision did not require semantic and phonological disambiguation of unvowelised words, although these two processes were running in parallel with the decision process.

When homographs are presented without a context that provides clues, only the dominant meaning is said to be accessed (Peleg et al., 2007). However, Rodd et al. (2016) have shown in an experiment involving members of a rowing club that specific experience can lead to individuals retrieving meanings other than the dominant one. Bentin and Frost (1987) proposed that, in Hebrew, the basis for lexical decisions for unvowelised homographs could be information that is shared by all lexical alternatives which are represented by the same consonant string. In other words, an abstract orthographic representation could provide enough information to enable a decision to be made. They proposed that this could occur in two stages, as suggested by Balota and Chumbley (1984); orthographic representation facilitates lexical access but decision is deferred until a phonological alternative has been selected. They also suggested that a lexical decision could initially be based on the shared but phonologically ambiguous letter string, followed by lexical disambiguation based on the higher frequency word being top-down generated and checked against bottom-up processing of vowels, with further options being generated in a cascade only if there is no initial match at the two levels. In word naming, initial lexical access is followed by a race between the various phonemic and semantic representations which are valid completed forms of the unvowelised letter string. The outcome of the race depends largely on which of the valid alternatives has the highest frequency.

The proposal that abstract orthographic representation could be sufficient for lexical decision is supported by the findings of Taouk and Coltheart (2004) that skilled readers are able to directly recognise words in their unvowelised form rather than recognising them via phonological forms. According to Abu-Rabia et al. (2003), a reading strategy that depends primarily on visualisation of orthographic features arises from extensive acquired orthographic knowledge. Ibrahim (2013) took this one stage further by proposing that skilled readers' preference for unvowelised consonant strings and visual-orthographic processing information pointed towards more accurate and faster reading of unvowelised words.

Studies of ambiguous words in other languages have provided more insights into what may be happening. Whilst in Hebrew and Arabic, ambiguity may be related to different pronunciations of an unvowelised letter string or to the same pronunciation having different meanings, in English, ambiguity is typically related to meaning; more than 80% of frequently encountered words have multiple meanings (Rodd et al., 2002). Despite considerable research into ambiguous words with multiple related senses (polysemous words), some scholars have perceived a shortage of clear models of how such words are represented or the processes involved in understanding them, and have proposed additional explanations such as semantic settling dynamics (SSD) within a connectionist network (Armstrong & Plaut, 2016). Pexman (2012) noted the apparent continuing lack of a model of the multiple dimensions of semantic information which could assist in understanding the processes.

Effects of ambiguity have been relatively well documented, although effects have varied depending on the task (Hino, Lupker, & Pexman, 2002). For example, response times in a lexical decision task are faster for ambiguous words with many related senses than those with multiple unrelated senses in various languages, such as English (Rodd et al., 2002) and Japanese (Hino et al., 2002). Two of the alternative explanations offered for this effect are a feedback account (Lupker, 2007), and attractor basins (Rodd, Gaskell, & Marslen-Wilson, 2004). The feedback account states that activation of an orthographic representation quickly leads to the semantic activation as well the phonological activation, after which the semantic feeds back to the orthographic representation and so on, with the number of multiple meanings creating stronger activation than occurs for words with only one meaning. This would explain the faster responses seen in lexical decision tasks but does not distinguish between related and unrelated meanings. The concept of attractor basins is based on representations reaching a stable (attractor) state. It assumes that unrelated meanings will create deeper basins than related meanings which will be located closely together and so form shallower and wider basins. The network will be able to enter the shallower basins more quickly than the deep ones. In this model, words with many related meanings would be recognised faster than those with unrelated meanings. Both the feedback account and attractor basins are assumed to operate within a connectionist model.

3.5 Psycholinguistic studies in Semitic languages

This section reviews selected psycholinguistic studies relevant to the present study. Studies conducted on reading in different languages have been broadly directed towards similar areas of interest. For example, many Arabic studies to date have focused primarily on the effect of vowelisation on accuracy of reading aloud and reading comprehension, especially in children, including a series of experiments by Abu-Rabia (see Taha, 2013), with a second and more recent strand of interest related to morphology (e.g. Boudelaa, & Marslen-Wilson, 2001, 2004, 2005, 2011; Boudelaa, Marslen-Wilson, Pulvermüller, Hauk, & Shtyrov, 2010). Hebrew studies have also addressed the impact of vowelisation (see e.g., Ibrahim 2013; Schiff, 2012) and aspects of morphological processing (see e.g. Deutsch, & Frost, 2003; Frost, Grainger, & Rastle, 2005).

3.5.1 Studies in Arabic

One of the disadvantages facing psycholinguistic studies in Arabic is that, until very recently, there has been no readily accessible corpus. A corpus allows variables to be counted and measured, which in turn enables researchers to control for the variables in experiments and hence contributes to the production of more rigorous research and reliable results. The corpus needs to be large enough and drawn from suitable sources to reflect the exposure of the general population to the language in order to establish representative statistics and norms, without which experiments have a baseline and results that can be difficult to justify. The lack of a sufficiently large and reliable corpus before Aralex (Boudelaa & Marslen-Wilson, 2010) has hindered research in Arabic psycholinguistics. Key pieces of research relevant to the present study are grouped under headings of influence of vowels, influence of consonants, morphology, and homographs.

3.5.1.1 The role and influence of vowels in Arabic

The impact of unvowelised printed text on reading Arabic has engaged researchers' attention for many years, although more from an educational than a psycholinguistic perspective. For example, a series of early studies focused on the effect of vowelisation on reading accuracy of single words, sentences and text, without reference to response times or reading speed (e.g. Abu-Rabia, 1995, 1996, 1997a, 1997b). Different age groups, reading skill levels and stimulus materials were introduced as the series progressed. Error rates in pronunciation were used to compare reading aloud from vowelised and unvowelised lists of words, sentences and text. Measured in this way, vowelised words and texts elicited fewer errors, leading to the conclusion that vowels were essential even for skilled adult readers. However, the conclusions drawn regarding theory from these studies can be critiqued on the grounds that the methodology did not lead to the conclusions. The methodology itself raises questions about the understanding of tasks; notably, a study using word naming of vowelised and unvowelised words included dismissal of the study by Frost et al. (1987) on the grounds that word naming was unsuitable for a Semitic language because it ignored the homograph phenomenon (Abu-Rabia, 1997b). The same study proposed that a connectionist approach involving the interaction of vowels and context could explain the role of vowels and context; phonological information provided by vowels was retained longer in working memory, and the context provided semantic priming. Vowels and context disambiguate meaning (Abu-Rabia, 1997b). It would appear that important aspects of theory have been lost in translation into English, except that similar confusion is evident in the description of the process by which skilled Arabic readers read unvowelised text as *"an interactive-dynamic process of context and word recognition, namely, if the Arabic text is presented unvowelized the only way to ensure word recognition is to rely on context"* (Abu-Rabia, 1997b, p.480). It is difficult to identify in these early studies some of the important details of the methodology used; for example, a study involving 10th grade Palestinian children with different skill levels refers to reading-disabled and poor readers as one and the same, when reading-disabled could include conditions such as dyslexia. In the same study, the skills judgment was based on a list of 70 words produced specifically for the experiment with no explanation of how the words were selected (Abu-Rabia, 1997b).

A further study, involving 78 Arabic speakers from 10th grade, used a range of materials again taken from curriculum textbooks (words, sentences and paragraphs). The tasks involved reading aloud 70 single words, 60 sentences and 15 paragraphs. These were vowelised, partially vowelised by adding the final vowel in each word, and unvowelised. One point was scored for each word read correctly, giving maximum points of 250 for paragraphs, 80 for sentences, and 70 for words. Skilled readers percentages correct were 77.6% for unvowelised paragraphs, 70% for sentences unvowelised and 13.4% for unvowelised isolated words. These results showed that, for skilled readers, overall accuracy was higher when more context was available. The only significant effect of vowelisation was in reading single words (Abu-Rabia, 1997b). Substantiating evidence that reading unvowelised words relies on activation of knowledge of morphology and vocabulary was provided by Frost (1994) and Abu-Rabia et al. (2003).

Research interest in the role of vowels in Arabic has continued (Abu-Rabia, 2001, 2002, 2012; Ibrahim, 2013), while further strands have developed regarding the role of consonants and morphology, and the acquisition of literacy.

3.5.1.2 The role and influence of consonants in Arabic

It is argued that unvowelised Arabic words cannot be correctly identified unless the reader can identify the missing vowels, a process which requires the identification of phonemes; phonemes can only be accessed by first identifying the morphemes, the intertwined consonantal root and pattern (Saiegh-Haddad & Geva, 2008). It is also proposed that the word-pattern, as an essential element of the word structure, facilitates access to the required information about the phonology. This information may be needed for correct pronunciation in word naming experiments.

It is also argued that children progress from reading fully vowelised text using phonological-recoding to encoding the whole words (Azzam, 1993). Azzam investigated reading errors and spelling errors in 150 native Arabic-speaking children aged from 6 to 11 years. She reported that reading errors mainly arose from the misreading of diacritics, which led them to pronounce an incorrect short vowel, whereas spelling errors were largely due to leaving out essential letters. She observed that children in Grade Two recognised words by their main graphic features, but children in Grade Three had mostly moved on to using letter-to-sound correspondences. Children in Grade Six still had problems with reading and spelling; diacritics were still misread, and incorrect short vowels pronounced. This study provided evidence that mastery was not achieved of the different phases of phonological recoding and orthographic realisation of whole words. It may also indicate that transitions between the stages were not sufficiently supported in the educational process. Taouk and Coltheart (2004) have similarly argued that whereas children read based on letter-sound correspondence, adults read by direct recognition of the unvowelised spellings, assisted by knowledge of any inflections and by the context.

The traditional way of teaching children to read in Arabic is to begin by teaching them the names of the letters, and next teaching them the three possible pronunciations of each letter together with the different ways of writing the letter depending on its position within the word, together with the relevant diacritics (Taouk & Coltheart, 2004). Books in use in preschools traditionally emphasized the sounds of consonants and the different ways of writing long vowels (Al Jarf, 2007). The first grade in primary school covers the same ground before teaching children to link letters together to read small words, with extensive repetition, rote learning and reading aloud. In the second grade, children are taught to read aloud small sections of text, consisting of approximately 2-4 lines. These reading skills are practised on text which progressively includes more lines, until children start to read texts which are only partially diacritised in the 4th grade. Reading aloud single words and building up to reading several lines is considered an essential component of gaining the phonological knowledge that will be essential later for reading unvowelised text. There is increasing reliance on recognising patterns which have previously been encountered, or guessing new ones, based on previous encounters with similar patterns. Many of these patterns will consist mainly or wholly of consonants (the root together with any long vowels). From the 5th grade onwards, it is a case of learning by experience and remembering what is learned.

A similar position exists for children learning Hebrew who also have to progress to reading unvowelised, highly homographic text as adults (Share, 2008). A triplex model of reading development has been proposed, with an initial stage of spelling-to-sound matching in Grade One, followed by a stage of processing strings (involving development of connections between morphological, lexical and orthographic knowledge) in Grade 2, and thereafter developing a contextual approach to reading (Share and Bar-On, 2017). Strengthening the development of links between orthography, morphology and lexical knowledge should assist the achievement of reading without short vowels.

Skilled adult readers, who have spent several or many years reading unvowelised text, may prefer to use the information they can see in the available letters and find it easier to read more accurately and quickly than the vowelised alternative (Ibrahim, 2013). Their response may also be one of automaticity. Boudelaa and Marslen-Wilson (2015) proposed that separation of roots composed of consonants and word patterns containing vowels could lead to differential processing in skilled readers, and through several series of experiments found that consonants acted as a prime when vowels did not. One series of experiments varied orthographic, morphological (root and word pattern), and semantic relationships between prime and target across four stimulus onset asynchronies (SOAs). Root effects were strong at all SOAs (32, 48, 64, and 80 ms), whereas word pattern effects were not observed at 32ms and, when seen, were short-lived (Boudelaa & Marslen-Wilson, 2005). Further experiments using priming identified that consonantal roots with a large family size elicited a faster response when the root was semantically transparent, 51 ms for the larger family size and 65 ms for the smaller (Boudelaa & Marslen-Wilson, 2011). When prime and target had different meanings but had the same morphological root, the priming effect was similar, although interestingly the family size made almost no difference, smaller root size eliciting responses in 51ms compared to 50ms for the larger family size. Auditory consonantal root primes were observed to exert the same influence as orthographic ones, indicating that the salience of the consonantal root is not lodged in the abjad alone (Boudelaa & Marslen-Wilson, 2011). Clearly there is also a close relationship between consonants and morphology in Arabic.

3.5.1.3 The role and influence of morphology

Printed Arabic text without short vowels contains words which look identical but have different meanings or senses which can only be correctly distinguished when the context is taken into account (Abu-Rabia, 2007; Saiegh-Haddad, 2004). Examples of homographs are, in English, 'Lead' (the metal) and 'Lead' (to guide) and, in Arabic, the word 'Qadam' قدم which can mean a foot as a part of the body or a unit of length.

A number of studies have reported that the complexity of Arabic orthography slows the word identification process, including Ibrahim et al. (2002) who reported that in a Trail Making Test, Arabic students responded 30ms more slowly than Hebrew students when the Arabic letters were connected, but responded almost as fast as the Hebrew students when the Arabic letters were not connected. The process of reading acquisition in the Arabic language is slower than it is in Hebrew (Abdelhadi, Eviatar & Ibrahim, 2011; Eviatar & Ibrahim, 2004). Based on comparative literacy rates, Myhill (2014) identified that most Arabic-speaking nations have lower literacy rates than expected from levels of income and investment in education. He attributed this to the fact that MSA was not a language that native speakers used in everyday life and proposed that children should receive education in their colloquial dialect in kindergarten and primary Grade One. A different view was taken by Saiegh-Haddad (2003) who attributed the problem to the phonemic distance between colloquial and formal Arabic and recommended early exposure to MSA, for example in stories.

Because Arabic is hypothesized to be a root-based language, it is proposed that morphological awareness plays a more potent role in Arabic than in English (Saiegh-Haddad & Geva, 2008). In Hebrew, experiments comparing priming by different classes of root with missing or silent items showed persisting influence of morphological constraints (Velan, Frost, & Deutsch, 2005). Morphological awareness has been linked with reading ability. For example, Abu-Rabia et al. (2003) found a significant difference in morphological awareness between two groups of Arab children, one group with, and the other without, reading disabilities. A later study found that the morphology and spelling measures were good indicators of reading accuracy and comprehension (Abu-Rabia, 2007).

Several studies have provided particularly compelling evidence for the reality of morphological structure (e.g. Boudelaa & Marslen-Wilson, 2001, 2004a, 2004b, 2005). However, there is no overall consensus on which elements of Arabic morphology are important or how they are represented in the mental lexicon.

Boudelaa and Marslen-Wilson (2001) examined whether the tri-consonantal root or an entity termed the etymon (a two-letter morpheme) was more likely to produce morphological priming effects. They provide the following example of an etymon where the core meaning is contained in two rather than three consonants. In the form ‘batara’(cutting) the core meaning is carried by the etymon ‘b,t’ which is found in related forms such as ‘batta’ and ‘batala’ (‘sever’ and ‘cut off’ respectively); not all the consonants of the root morpheme ‘b,t,r’ are present (Boudelaa & Marslen-Wilson, 2001, p.65). The etymon was proposed as a way of explaining why bi-consonantal roots existed, and explaining phonetic and semantic relationships between words, based on observations that some words could have the same root but no noticeable shared semantics, while other words could have the same root yet have very different meanings (Bohas, 2006). Boudelaa and Marslen-Wilson (2001) conducted two experiments with 30 volunteers, Tunisian high school students aged from 16 to 20 who used MSA in their studies. The two experiments were cross-modal immediate repetition priming and masked morphological priming Results indicated that in both experiments priming occurred with the etymon, whether the prime and target were both morphologically and semantically related or simply morphologically related. They concluded that the etymon was a finer-grained unit than the tri-consonantal root which could also function as an organizing unit in the MSA lexicon. This led them to suggest that interactive models would need an additional layer between the letter and the morpheme, and to propose that connectionist models could accommodate additional relationships between etymon and triconsonantal root through the productivity of the root

A 2004 study by Abu-Rabia and Awwad hypothesized that Arabic readers would access meaning via the root but use grammatical clues in the affixes and suffixes, for example pronouns, to work out the most appropriate word. They tested this with a group of 48 skilled readers aged between 16 and 18, using four experiments with masked priming, lexical decision and naming based on noun derivation. The results showed that priming with a word which had the same word pattern did not significantly facilitate either naming or lexical decision. This led the authors to conclude that organisation of the mental lexicon was not dependent on the principal morphemic units of roots and word patterns, and that whole words were recognised rather than their components. In particular, this research focused on the function of morphological components of noun derivations with regard to lexical access. Nouns with their morphological patterns were found to be recognized as whole words and thus were represented as complete shapes in the mental lexicon, each one with a separate representation (Abu-Rabia, & Awwad, 2004).

Two further aspects of Arabic morphology were studied by Boudelaa and Marslen-Wilson (2004a, 2004b): allomorphs and the lexical representation of abstract morphemes in the CV-skeleton (the arrangement of consonants and vowels which determines all the phonemes of the complete word). Allomorphic variation (Boudelaa and Marslen-Wilson, 2004a) occurs when the meaning of a morpheme is unchanged but the sound is different. For example, in English, the bound morpheme *'ed'* denoting the past tense has three variant sounds: [d] as is 'played', [t] as is 'walked', and [id] as in 'limited'. All stimuli were unvowelised and either unambiguous or had a strongly dominant meaning. The first experiment used the cross-modal priming lexical decision task (with the spoken prime immediately followed by a visual target which requires lexical decision) with strong roots that did not change as well as weak roots in which not only did sound change but also only two of the three root letters appeared. Results indicated that priming was consistent for word pairs which had the same root morpheme, whether or not the root was strong, provided the weak root appeared in tri-consonantal form in the target even if its semantic meaning was different. A second experiment investigated how allomorphy affected processing of word patterns by comparing different types of word pairs. In some pairs the word pattern was transparent in both prime and target, in others a weak root caused assimilation process in the prime without altering the CV-skeleton structure of the word pattern, and in a third group of pairs the CV-skeleton structure was disrupted in prime and target alike because the glide of the root was allomorphically deleted, changing the number of vowels and consonants from the number in the underlying form. Strong priming effects were observed in the first two conditions but not in the third, leading to the conclusion that a change in surface phonology did not affect the role of the consonantal root in the lexical access process. Moreover, the absence of priming when the CV-skeleton was disrupted implied that access to representations in the mental lexicon was also disrupted, suggesting an important role for the CV-skeleton. Even though the word pairs had no consonants or vowels in common, the CV-skeleton they shared was found to speed lexical decision (Boudelaa & Marslen-Wilson, 2004a). This study provided further evidence that lexical decision in Arabic was influenced by morphological features, in this case the abstract CV-skeleton. Frost et al. (1997) proposed a dual-route model which could explain the role of morphology in processing Hebrew words. This model assumed a level of words and a subword level consisting of root morphemes. These were connected so that the root could be accessed directly by a process of morphological decomposition or from the printed word which contained the root. Frost hypothesised that these processes acted in parallel. The findings of Boudelaa and Marslen-Wilson (2004b) suggest that there may be more than one subword level involved in the processing of words in Semitic languages. Alternatively, connectionist models assume that morphological structures are not part of the mental lexicon. Instead, as the network learns patterns of activity which are based on mappings of phonological and orthographic repeated features and their associated meanings, the morphological patterns will be captured because of regular correspondences.

One study, involving 30 volunteers who were skilled high school students aged 16 to 20, used a total of 96 target verb forms without diacritics to investigate whether the abstract linguistic unit of the CV-Skeleton functions as a morpheme in the psycholinguistic, cognitive domain and how morphological structure influences cognitive processing (Boudelaa & Marslen-Wilson (2004b). For this study, 96 pseudowords were created by combining a non-existent consonantal morpheme with a real skeletal morpheme and vocalic morpheme. The first masked priming experiment was used to prime between morphologically related words that shared a skeletal morpheme, a vocalic morpheme, or both. Masked priming was chosen because it has been considered to be insensitive to semantic effects while facilitating the study of morphological and form-based effects (Rastle et al., 2000). The results of this experiment indicated that priming by CV-skeleton occurred when the skeleton, consonant and vowel sequence, and sequence-associated syntactic meaning were the only shared elements of prime and target. This suggested that the CV-skeleton could play a significant role in mapping orthographic forms onto lexical representations, although findings were not statistically significant. A further 36 volunteers from the same high school took part in the second, cross modal priming experiment. This presented an auditory prime with a fully articulated vocalic morpheme in order to specifically examine the role of that morpheme. They also investigated whether the role of the essentially phonological CV-skeleton in processing surface forms was stronger if the prime was spoken. Results showed strong priming effects for both CV-skeleton and word pattern, but not for the vocalic melody, suggesting that the CV-skeleton functions as an abstract representation. The third experiment used auditory primes and targets, and findings confirmed facilitation by a skeleton morpheme, or a skeleton and vocalic morpheme together, but there were no significant effects for the vocalic morpheme. It was concluded that the CV-skeleton is a distinct lexical entity.

A further study investigated the activation of different morphemes over time in a lexical decision task in which 24 prime-target pairs of productive (deverbal) and non-productive nouns without short vowels were assigned to each of six conditions (Boudelaa and Marslen-Wilson, 2005). Conditions were: shared word pattern with average 1.2 letters shared; average 0.9 letters the same but unrelated semantically or morphologically (orthographic condition); shared root and semantic interpretation; shared root but different semantic interpretation; shared root and average 2.5 letters shared (orthographic condition); and semantic relationship but not a morphological one. Results showed different patterns for effects of morphology as compared with semantic and orthographic conditions; although orthography showed a short-lived early priming effect, priming by a shared word pattern had a significant facilitatory effect. Moreover, the effect of priming by roots not only appeared very early but was maintained. It was proposed that the root might play a major role in visual word recognition in Arabic. Due to a later and shorter priming effect of word patterns, it was further proposed that word patterns were identified as separate morphological units. Moreover, morphological effects appear to take precedence over effects of form and meaning. Effects of orthographic and semantic priming appeared later than the morphological effects. The finding of Abu-Rabia (2012) that initial lexical access is via the root is consistent with these findings from Boudelaa and Marslen-Wilson (2005) since priming by roots has an earlier and longer-lasting effect than priming by word patterns. Priming by deverbal noun word patterns and verb word patterns suggested they would need to be included as entries in the Arabic mental lexicon as well as roots (Boudelaa & Marslen-Wilson, 2005).

3.5.2 Comparisons of Arabic and other languages

Some studies have highlighted differences between results in Arabic and Hebrew, which point to differences between how the mental lexicon is organised and accessed. Frost (1994) identified different effects of semantic priming for vowelised and unvowelised Hebrew. Masked priming experiments investigating priming by nominal and verbal word patterns in Arabic found both forms to be equally robust (Boudelaa & Marslen-Wilson, 2001, 2005), whereas priming experiments in Hebrew found a lack of priming in Hebrew nouns which had a common word pattern, in both masked and cross-modal priming (Frost, Deutsch, Gilboa, Tannebaum, & Marslen-Wilson, 2000; Frost, Forster, & Deutsch, 1997). This led Boudelaa and Marslen-Wilson (2011) to advise caution regarding attempts to link Arabic and Hebrew data.

Frost et al. (1987) compared strategies for visual word recognition in three languages with varying orthographic depth. 48 native speaker students of Hebrew, English and Serbo-Croatian were drawn from each of three universities. In each language, 24 students were randomly allocated to a naming task and 24 to a lexical decision task. Both tasks involved the same list of 48 non-words and 48 words; homographs in Hebrew were excluded. Half the words were low frequency and half high frequency, determined by averaging ratings given by different native speakers on a 5-point scale; all were 3-7 letters and 4-6 phonemes in length, with the average length of words in Hebrew being shorter due to the omission of vowels. Non-words were generated by changing a single letter of a real word. Results indicated that word naming was slower than lexical decision in Hebrew, but only for high-frequency words, whereas word naming was faster than lexical decision in English and Serbo-Croatian words. High frequency words were named faster than low frequency words in all languages but lexical status (non-word, high- or low-frequency word) had a greater effect than frequency, with a bigger impact in Hebrew than English. It was concluded that naming in Hebrew was lexically facilitated and could not be achieved until a lexical decision had been made. However, it was noted that increasing the proportion of non-words in a naming list from 20% to 80% reduced the response time and may have implied a change of strategy to using the nonlexical route because non-words do not exist in the lexicon. The more usual reliance on the lexical route due to the absence of short vowels may also apply to Arabic.

In Hebrew, one early study employed semantic priming to investigate disambiguation of heterophonic and homophonic homographs (Frost & Bentin, 1992). Ambiguous primes preceded unambiguous targets, and stimulus onset asynchrony (SOA) was set at 100 msec, 250 msec, and 750 msec. It was found that for heterophonic homograph primes, the facilitation of lexical decision related to the subordinate phonological alternatives occurred only at 250 and 750 msec, in contrast to the dominant phonological alternative which occurred at all SOAs. This was considered to support the ordered access of related meanings. For homophonic homograph primes, lexical decision was facilitated at all SOAs, irrespective of the dominance of meaning. Four explanations were given for these results: access to meaning mediated by phonology; multiple lexical entries for heterophonic homographs; single lexical entries for homophonic homographs; and characteristics specific to the Hebrew language responsible for the long duration of subordinate meaning activation (Frost & Bentin, 1992). These explanations suggest that in the present study, responses in the word naming task will be slower if there are multiple lexical entries for heterophonic homographs, but that it will not be possible to distinguish between the effects of heterophonic and homophonic homographs.

A further study (Peleg, Markus, & Eviatar, 2012) investigated how and when heterophonic and homophonic homographs in Hebrew were processed by the brain. Participants were required to read homographs following a sentence context, biasing or non-biasing, and to perform a lexical decision task on targets presented laterally 1000 ms after the onset of the ambiguous prime which ended the sentence. Targets were variously unrelated and related to either the subordinate or dominant meaning of the homograph. Results indicated that activation of dominant and appropriate meanings was both quicker and more durable than subordinate and inappropriate meanings. However, neutral contexts were associated with retention of the dominant meaning alone, whereas biased contexts were associated with retention of the appropriate meaning alone. The present study involving isolated letter strings could therefore be expected to relate to the dominant meaning, which requires identification before use in word naming experiments.

3.6 A note on diglossia

Diglossia refers to a situation in which two languages (or variants of the same language) are used within a community, often by the same speakers but under different conditions, typically in colloquial, informal settings and in formal settings in education and public life. Along with the complex orthography of written Arabic, it has been proposed that the unique sociolinguistic properties arising from diglossia may affect reading and the acquisition of literacy. There are many different versions of colloquial Arabic depending on the country in question. Thus, the situation regarding the effect of diglossia may vary among countries in which MSA is the official language. Some scholars think that diglossia has an effect on the acquisition of basic academic skills in Arabic during the early years of education (e.g. Ayari, 1996). There are differences between colloquial and Modern Standard Arabic regarding the vocabulary, pronunciation, syntax, and grammar, which means that the language in which children read is not the same as the one they generally use with family and other social contacts. Ayari (1996) has argued that Arabic diglossia results in difficulties in the acquisition of reading skill. He attributed this to differences in pronunciation between colloquial Arabic and MSA which may imply two separate phonological lexicons. Many Arabic children are not exposed to literary Arabic until the first grade, when it is similar to a second language. It has also been suggested that the phonological distance between the first language and MSA may play a part in visual word recognition which relies on non-phonological resources but requires matching between sounds and letters (Saiegh-Haddad 2005; Saiegh-Haddad, & Geva 2008). Children have to learn unfamiliar phonemes and match those to MSA letters without being able to match many of the words and sounds they already know to print.

One study examining the impact on reading acquisition of early exposure to literary Arabic found that children exposed to literary Arabic in the pre-school stage achieved higher reading comprehension scores at the end of grades 1 and 2 than children who had been exposed only to spoken Arabic (Abu-Rabia, 2000). These findings confirmed those of Iraqi (1990) who investigated the effect of daily reading of stories in literary Arabic to preschool age children on their oral skills as well as their listening comprehension. The experimental group who listened to stories in literary Arabic for 15-20 minutes daily for five months demonstrated higher levels of oral ability as well as listening comprehension compared with a control group who had been exposed to daily stories in spoken Arabic. This may be related to the acquisition of vocabulary.

Vocabulary is considered a prerequisite for success in reading comprehension (Joshi, 2005; Salah, 2008). It has been estimated that, in English, a vocabulary of 5,000 or above is needed for normal everyday reading; below that level, which represents some 95% of the words likely to be encountered, reading comprehension is considered to be inadequate. The comparable figures for Arabic generated by one study were a vocabulary size of 6,000 which represented was 90% of the words likely to be encountered in Arabic (Salah, 2008). It is interesting that the percentage for Arabic is lower than for English. Although Salah was concerned with the large size of vocabulary required, the lower percentage may also indicate the role of deduction of meaning in reading an Arabic text.

It has been shown in some studies that MSA and colloquial Arabic act as separate language systems in the brain (e.g., Ibrahim & Aharon-Peretz, 2005). This implies that literate Arabic speakers may be bilingual, assuming that they have achieved a sufficient level of spoken fluency in both. Ibrahim, Eviatar and Aharon-Peretz (2007) conducted a comparison of reading measures in children in the first grade; some were Arab monolingual, some Hebrew monolingual, and others were bilingual in Hebrew-Russian. Measures were accuracy and rate of reading text, pronunciation errors in reading single and non-words, and correct answers to the phonological awareness test. This test asked children to identify the initial sound (not syllable) of each of 20 words spoken by the researcher, to identify the final sound (not syllable) of each of 20 words spoken by the researcher, and to pronounce 20 words but to delete one the syllables in the words. Despite the fact that Arabic native speakers achieved higher scores than Hebrew monolinguals in terms of their phonological awareness, this did not help them to read more easily; they read more quickly but made more errors. The researchers concluded that the native Arabic speakers experienced more difficulty than Hebrew monolinguals and bilinguals in language processing, which they proposed might be related to the visual complexity of Arabic orthography.

Bentin and Ibrahim (1996) found that the reaction times for recognising printed Arabic words by Arabic speakers were longer than those for recognition of Hebrew words by Hebrew speakers and hypothesized that this reflected the fact that diglossic Arabic speakers were less familiar with the written language. This was partly based on an unusually large frequency effect which was three times larger in printed words than in spoken ones and partly on slower responses to transliterations of their spoken Palestinian dialect, even in a delayed naming task. A separate study reported that response times for printed Hebrew stimuli were more rapid than those for Arabic word recognition (Ibrahim, Eviatar, & Aharon-Peretz, 2002). Eviatar and Ibrahim (2004) investigated the initial stage of learning to read Arabic using two tasks (a lateralized (CVC) identification task and a letter matching task) among Arabic, Hebrew and English speakers. English speakers required the shortest exposure duration for letter identification, Arabic speakers required the longest duration, and Hebrew speakers required an exposure duration between the two. These findings provide additional evidence of the complexity of word recognition in Arabic.

3.7 Implications for the organisation of the Arabic mental lexicon

Diglossia raises interesting but unanswered questions about the organisation of the Arabic mental lexicon: whether colloquial Arabic in daily use but without a written form has a ‘sound dictionary’ for the meanings of words, and whether MSA later develops as a separate part of the mental lexicon or as some form of extension to the first (a subject of continuing debate regarding bilinguals and second language learners). Development of the mental lexicon through the years of education progresses through letter names, phonemes, fully diacritised words, and partially diacritised words to using unvowelised printed text. The importance of morphological aspects of organisation has been highlighted throughout the second chapter on the characteristics of the Arabic language as well as the discussion of psycholinguistics studies in the present chapter. However, this has led to an emphasis in research on morphological structure in the mental lexicon rather than more general considerations.

Some of the studies reported in this chapter have found that adult native speakers take longer to name vowelised letter strings than those without short vowels, which may indicate that, for familiar words at least, there is a more immediate association between an unvowelised letter string and the whole word. For children who are beginning readers this is not the case; they rely on the full orthographic form of the word. From a purely logical point of view rather than a psycholinguistic one, this may imply that there are various ways in which words can be represented in the lexicon. Complete high frequency MSA words could be represented by unvowelised letter strings, which in this context could be considered ‘words’. Less frequent and unfamiliar words appear to require morphological decomposition and reassembly to reach the full orthographic and phonemic representations of the word. This is achieved through some combination of roots, word patterns and the CV-skeleton. There is no evidence to date to support further speculation about how the Arabic mental lexicon is organised (Boudelaa & Marslen-Wilson, 2015). Exploration of how heterophonic homographs are processed in the word naming and lexical decision tasks covered in chapters six and seven in the present may help to shed light on the organisation of the lexicon.

3.8 Word naming and lexical decision tasks

Two of the most widely employed tasks for studying the cognitive processes involved in visual recognition of the printed word are word naming and lexical decision (Katz et al., 2011). Word naming requires participants to pronounce aloud, as quickly as they can, a printed letter string they see on a computer screen. Lexical decision requires them to decide, again as quickly as they can, whether a printed letter string they see on a computer screen is a word or not; they are asked to press one button if the string is a word and another button if the string is a nonword. Asking participants to respond quickly is considered to make the laboratory tasks more closely resemble reading in real life. Response times are frequently used as on-line measures since they give access to the cognitive processes involved. In word naming, specialist software is used to measure and record in milliseconds when a participant starts to pronounce the letter string they see on screen. The researcher may need to use additional software in order to listen carefully to each response to ensure that the time recorded is accurate and does not reflect any sounds made before the start of the pronunciation such as the noise made by lips opening, as noted by Protopapas (2007). In lexical decision, measurement in milliseconds of the time taken to press a button or key is automatically recorded by the software. Participants are given a number of practice trials before starting the experimental trial; target items used in the practice trial should be typical of those in the experiment but not be included in the experiment itself.

The tasks are used to examine the effect of diverse variables on visual word identification so that their relative influence can be evaluated, and insights gained into the cognitive processes involved. In order to identify the effect of a particular variable, it is important to know and control for other variables by careful selection of stimuli. For example, in the present study it is important to know which letter strings are understood as homographs by participants so that letter strings are correctly classified. It is essential to know the frequency and other characteristics of target items so that any differences in speed of responses can be correctly attributed to the variables which cause them. Word frequency (section 3.2.1) affects the speed of participant response, with high-frequency words being recognised more quickly. The present study uses frequency of each letter string as it appears in unvowelised printed text in the Aralex corpus (Boudelaa & Marslen-Wilson, 2010). Word lengths (section 3.2.2) can be measured in letters, phonemes or syllables; for the present study, length in letters is the appropriate measure because short vowels are omitted. This enables the effect of length of a letter string, the visual stimulus, to be identified and accounted for, since shorter words elicit faster responses. Response times increase with more letters. For a study of homographs, it is necessary to be able to measure the extent of homography (section 3.2.4) in order to compare the speed of response to letter strings with more or fewer pronounceable forms because that may shed light on how they are processed, and whether they are responded to more quickly or more slowly. In the present study it is also important to include the effect of AoA (section 3.2.5) because learning MSA on starting school with its different vocabulary and phonology may affect responses, and the spoken dialect used before starting school will not have exposed children to heterophonic homographs. It may be necessary to separate the effect of frequency from the effect of AoA because that may provide clues about a development model of reading.

Problems can arise in the lexical decision task when there is no control for potentially influential variables such as frequency, word length, and neighbourhood density, problems which it is asserted can be revealed by the particular characteristics of just one set of 10 to 20 stimuli and which restrict the possibility of generalising the findings (Balota et al., 2007). Words therefore have to be matched on a range of variables to try to ensure that it is indeed the target variables that are being measured. For example, in a lexical decision task, many variables have been shown to affect lexical access of words and non-words. The variables concerned include letter and phoneme length, orthographic similarity or regularity, and regularity of similar words in addition to word frequency (see e.g., Cutler & Davis, 2012). Further variables include Age-of-Acquisition (AoA), imageability, morphological transparency, n-gram or bigram frequency, and orthographic depth (Adelman, 2012).

Together with lexical decision tasks, word naming tasks have been the principal method for investigation of the role of phonology and associated decoding processes in printed word identification (e.g., Rastle & Coltheart, 2006). Like lexical decision tasks, research has shown that performance in word naming tasks is also influenced by many variables which can operate at varying levels, namely the lexical, semantic and superficial levels (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). Performance at the lexical level can be affected by variables such as the regularity of the orthographic system (González-Nosti, Barbón, Rodríguez-Ferreiro, & Cuetos, 2014), spelling-to-sound consistency and word length. At the semantic level, imageability has also been shown to exert an influence (Balota et al., 2004).

It has been recognized that it is insufficient to simply report that particular variables affected lexical processing: because the extent of an individual variable's contribution was also important (Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004). The contribution of a range of variables in three categories, namely lexical, phonological and surface level study was investigated by Balota et al. (2004). In their study, the selected lexical variables included word length, subjective frequency (familiarity) and orthographic neighbourhood size. For semantic variables, imageability, meaningfulness and connectivity were measured in lexical decision and speeded visual word naming tasks based on 2,428 single-syllable English words. Tasks were administered to a sample of thirty young adults and thirty older adults using regression techniques to analyse the data.

It was found that for most variables, their influence was task dependent, with the greatest semantic level influences being seen in the lexical decision task (Balota et al., 2004). Furthermore, care is needed in interpreting results which may be influenced by the additive effects of a target variable such as word frequency with the quality of the stimulus itself within the processes of lexical decision (Balota, Aschenbrenner & Yap, 2013). Thus, the importance of stimulus selection is highlighted.

The need for care in interpreting results also relates to the fact that the characteristics of individual participants within a group may distort results. For example, it is not known whether a smaller or larger vocabulary enhances or inhibits speed of response (Katz et al., 2012).

Care is also required in attributing results to particular causes; in recent years, the part played by morphology in visual word recognition in English has been investigated (Marslen-Wilson, Bozic, & Randall, 2008). One study consisted of two visual priming experiments with varying relationships between orthography, semantics and morphology under three different SOA conditions (Rastle et al. 2000). In English, significant effects were found for the influence of morphological structure on early visual recognition, which were distinct from any orthographic or semantic relatedness of the words. This led the authors to postulate that there are orthographic representations based on morphemic structure which are not determined by semantic transparency.

This is even more important for some of the very different orthographies in use in many other languages. In Semitic languages, for instance, some studies have found less influence of orthography on word recognition by skilled readers than in English, although morphology has been found to exert an influence (sections 2.3.1 and 3.5.1.3), in Hebrew (Deutsch, Frost, Pollatsek, & Rayner, 2000; Velan & Frost, 2011) and in Arabic (Balota et al., 2006; Boudelaa, & Marslen-Wilson, 2011).

In the search for explanations of the relationships between the phonological, semantic and written representations of words, it is asserted that the morphologies of Semitic languages are a rich resource for reading researchers (Balota et al., 2006).

3.8.1 Word naming

The word naming task requires participants to read aloud a series of given words as quickly as possible, one at a time as they are presented on screen, and responses are measured in terms of accuracy and speed of response. It is relatively straightforward because it is directly a reading task without the need to include non-words. In addition, it has been said to be a better test of automaticity because fewer processing strategies are involved; it avoids the stage of decision which is not found in real-life reading (Balota & Chumbley, 1984). It also has a production component, which involves phonology, thus placing greater reliance on the ability to correctly associate spelling with sound (Balota, Paul, & Spieler, 1999). In common with the lexical decision task, it has a lexical access component, but is more straightforward to administer. However, analysis of the sound files can be time-consuming.

AoA has been shown to be an influential variable in word naming task performance (Morrison & Ellis, 2000). Unlike earlier studies of AoA effects, which used adult estimates of the age at which words were learned, Morrison and Ellis used word naming with children of varying ages as the basis for deriving objective AoA norms.

While the word naming task is a better test of automaticity, the lexical decision task provides opportunities to investigate how decisions are made at the strategic level whether consciously or otherwise (Chumbley & Balota, 1984). Similarly, semantic effects of variables are typically greater in lexical decisions than in word naming (Cortese & Khanna, 2007), enabling different insights into how letter strings are processed.

3.8.2 Lexical decision

The lexical decision task (LDT) at its most straightforward involves presenting study participants with a succession of letter strings which appear on a computer screen and participants have to decide whether each string is a real word (e.g., clear) or a non-word (e.g., claer) (Jones, & Estes, 2012). Participants press one key to indicate ‘yes’ for a real word and ‘no’ for a non-word. If the non-word is orthographically legal and pronounceable, the decision has to be made by accessing the mental lexicon to find out if the word is known. Letter strings are typically presented one at a time, either a real word or a non-word.

It has been demonstrated that a number of variables affect performance in lexical decision tasks. These variables include: word frequency (Balota & Chumbley, 1984; Balota & Spieler, 1999), word length (New, Ferrand, Pallier, & Brysbaert, 2006), number of syllables (Ferrand & New, 2003), AoA (Cortese & Khanna, 2007), and orthographic regularity (Balota et al., 2006), among others. One particular feature of lexical decision tasks is the inclusion of non-words in order to achieve improved discrimination of correct answers arrived at by educated guesswork from those arrived at through a systematic process of lexical access.

Properties of both real words and non-words need to be established in order to meaningfully analyse and interpret data. For example, if non-words cannot be pronounced, this will impact the depth of processing (Jones & Estez, 2012). In addition, the percentage of non-words needs to be specified.

The way in which non-words are created can affect the results of a lexical decision task and it is therefore important to know how researchers have created their non-words. In the present study, the process of creating non-words is described in the chapter reporting observations from the lexical decision task (Chapter 7). Non-words can be categorized according to pronunciation and orthography; they can be homophones of real words, pronounceable and orthographically legal, or both unpronounceable and orthographically illegal (Yap, Balota, Cortese & Watson, 2006). An example of a homophone of a real word is *drane* (for *drain*), an example of a non-word which is pronounceable and orthographically legal (also termed a pseudoword) is *darne*, and an unpronounceable and orthographically illegal example is *dnraeI*. The degree to which the non-word resembles a real word has been shown to strongly affect lexical decision latencies. One study found that pseudowords in which one letter of a real word was transposed produced different latencies from pseudowords in which one letter of a real word was replaced (Perea, Rosa, & Gómez, 2005). This led to the proposition that the similarity between pseudowords with transposed letters and their associated real words was greater than the similarity between pseudowords with replaced letters and their associated real words. An associated frequency effect was identified (Perea, Rosa, & Gómez, 2005). Furthermore, interaction between non-word type and word frequency can strongly affect the word frequency effect (Yap et al., 2006). Thus, it is important to clearly establish attributes of both real words and non-words in order to meaningfully analysis and interpret data. For example, if non-words cannot be pronounced, this will impact the way in which they are processed. In Arabic, the most important feature to maintain pronounceability is to have a recognisable structure of consonantal but false root and a recognisable pattern of consonants and vowels, and to avoid simply inverting a letter or creating another real root and word when changing a root consonant. It has also been suggested that the inclusion of such words in studies which involved phonological recoding could cause subjects to process them in an atypical way (Baluch & Besner, 1991).

An important function of the results of these tasks is that, depending on experiment design and control of relevant variables, they can be used to test existing theories of world recognition and reading, and potentially add to our understanding of the processes in a range of languages.

3.9 Summary

This literature review has explored key themes and previous research relevant to an understanding of how orthographic and morphological attributes affect Arabic reading. The main characteristics of word naming and lexical decision tasks have been presented, followed by a discussion of some of the main theories of reading (interactive activation, dual route and connectionist) in addition to some later modifications of these theories. Then some of the key variables which can influence the outcome of experiments were examined, in particular frequency, word length, neighbourhood density and AoA. The ensuing discussion of disambiguation of homographs identified the characteristics of Arabic which result in a particularly high number of homographs and potential processes of disambiguation. Disambiguation of homographs may be achieved through visual pattern recognition, frequency, contextual factors or some other process. Further investigation into this area may confirm existing findings or shed new light on reading processes in Arabic.

A selection of studies into aspects of word recognition and reading was reviewed, indicating that, whether the skilled reader of a Semitic language is reading isolated words or sentences, different strategies and processes are employed. Whereas in reading aloud sentences, skilled readers can correctly deduce the meaning of heterophonic homographs from grammatical or semantic clues in the context, in reading aloud isolated words (particularly low frequency words) they may have to rely on morphological decomposition and reassembly to reach the pronunciation. As schoolchildren, they progress through several distinct strategies: reading aloud vowelised text, phoneme by phoneme; reading aloud and visual recognition of fully vowelised letter strings; reading aloud and visual recognition of partially vowelised letter strings increasing reliance on visual-orthographic information and contextual clues (Azzam, 1993: Share & Bar-On, 2017; Taouk & Coltheart, 2004). Adult native speakers of Hebrew use a similar range of strategies, which may apply in Arabic but requires further investigation. For example, in reading aloud isolated words, Hebrew speakers are considered to use serial phonological assembly for vowelised letter strings because response times are longer than for pronunciation of unvowelised letter strings which are considered to give direct access to pronunciation (Weiss et al., 2015). In addition, adults use a range of strategies related to morphological decomposition: identification and decomposition of morphemes (Deutsch et al., 1988; Verhoeven & Perfetti, 2011) and different processes for identification and pronunciation of morphemes in words derived from Semitic and Indo-European languages (Velan & Frost, 2011). The section concluded by looking at the phenomenon of diglossia in Arabic.

In sum, research reveals that the Arabic language is complex in nature and characteristics, hence the complex processes involved in word recognition and language comprehension merit further study. The results of experiments indicate that the orthographic characteristics of Arabic place additional visual and cognitive demands on readers. The reading process in Arabic for skilled readers cannot be viewed simply as a process of word recognition. The reading process for skilled readers reading unvowelised Arabic texts should be viewed as an interactive-dynamic process of context and word recognition, namely, if the Arabic text is presented unvowelised and without diacritics, the only way to ensure word recognition is to rely on context. However, children and adults can and do read single words thus studies were reported which implies a different set of processes.

The literature on the relationship between reading and language development in Arabic is limited in comparison with English and other European languages, despite the increased interest in investigating this relationship. Arabic and English differ in their orthographic depth and morphological transparency (Saiegh-Haddad & Geva, 2008). Studies involving languages with greater orthographic depth, such as

Arabic and Hebrew, offer the possibility to make informative comparisons of the cumulative as well as the unique contribution of language components to the reading process. The following chapter describes the development of a database of 1,474 Arabic words with the aim of enabling such investigations.

Chapter 4 A STANDARDIZED SET OF 1,474 UNVOWELISED ARABIC LETTER STRINGS: NAME AGREEMENT, FREQUENCY AND WORD LENGTH

4.1 Introduction

This chapter justifies the name agreement study and explains how it was conducted. Initial word selection and name agreement procedures are described, followed by statistical analysis of the survey results for participants and items, and presentation of the predictive variable models for the survey data. The chapter concludes with discussion of the survey findings.

There was no directly comparable study in Arabic published at the time of the present study. Available data regarding name agreement procedures was found in the development of pictorial stimuli databases in Tunisian Arabic (Boukadi, Zouaidi, & Wilson, 2016) and Jordanian Levantine Arabic (Khwaileh, Body, & Herbert, 2014) that cannot be directly compared with MSA.

Research into natural Arabic cognitive reading processes involved in reading unvowelised printed text is scarce and no research has been undertaken to establish which unvowelised letter strings are read by normal readers as homographs rather than as a single possibility. Previous studies involving homographs have typically used fewer than 100 unvowelised homographic letter strings (e.g. Bentin & Frost, 1987; Frost & Bentin, 1992; Seraye, 2004) and have mainly used dictionaries and experts to determine lists of pairs of higher frequency and lower frequency alternatives. It is important to have a sufficiently large set of unvowelised letter strings to try to ensure that the stimuli are sufficiently representative, and the results of statistical analysis are reliable. The analysis of name agreement may provide an insight into the operation of concepts of lexical selection and is therefore potentially useful to understanding how Arabic readers process heterophonic homographs. In order to consider the effects on reading of the homographic orthography, it is also important to involve as many participants as possible in determining which letter strings they read as homographs and how many alternatives they know because they may know only one or two of the possibilities listed in dictionaries. Differing levels of education and ages among participants also help to ensure that survey findings are representative of the wider population.

However, research involving homographs may require advance knowledge of the relative dominance of the different word forms or meanings (Gawlick-Grendell & Woltz, 1994). The usefulness of isolated heterophonic homographs in psycholinguistic experiments, where there is no biasing context, depends on the existence of meaning frequency norms. When homographs are presented without a context that provides clues, only the dominant meaning is accessed (Peleg et al., 2007). For isolated homographic words, some meanings have a greater probability of occurrence than others (Gee & Harris, 2010). Thus, it is important to know which meaning participants are most likely to think of first, the dominant meaning. In the case of heterophonic homographs in Arabic, the dominant meaning will be the same as the dominant vowelised orthographic form.

This study presents a standardized set of 1,474 unvowelised Arabic letter strings for use in psycholinguistic experiments investigating similarities and differences in processing unvowelised non-homographic and homographic letter strings. The set consists of 887 unvowelised letter strings which can be vowelised and pronounced in more than one way (heterophonic homographs) and 587 unvowelised letter strings which have only one possible form of vowelisation and pronunciation (non-homographic). They have been standardised on name agreement, which is the extent to which participants agree on a particular vowelised form of a letter string, and hence on its pronunciation and meaning. The present study follows the same data collection process which is employed in picture naming studies, but presents an unvowelised letter string in place of a picture. Like picture name agreement studies, the data of interest were how each participant chose to produce an output sound (a pronunciation), where multiple pronunciations were compatible with the presented stimulus. Variations in the pronunciations produced by individuals were also of interest. As with pictures, some stimuli would elicit responses with a higher level of agreement than others, and some would elicit more diverse responses.

Name agreement was determined by the highest proportion of participants identifying the same vowelised form of the letter string and was given by the modal response, in preference to setting an arbitrary percentage threshold. Name agreement can also be understood as meaning dominance, that is the meaning agreed on by more participants than any other. Meaning dominance norms were collected for 887 heterophonic homographs using normative data collected from a relatively large sample of participants (n=445). The length (in number of letters) and the frequency of the unvowelised letter string as extracted from Aralex (Boudelaa & Marslen-Wilson, 2010) are also provided in the database.

Name agreement was essential to confirm which words were understood by skilled Arabic readers to be homographs, and to determine the dominant vowelised form and meaning of the homograph in order to provide a set of standard responses against which responses in the word naming tasks could be measured. This also enabled the categorisation of words represented by homographic letter strings into nouns and verbs, resulting in a total for all letter strings of 1,064 nouns and 410 verbs within the final list. The potential importance of this normative variable to reading aloud and visual word recognition is discussed. An extract from the list of unvowelised letter strings is provided in Appendix A and an extract from the list of unvowelised non-word letter strings is given in Appendix B. The next section offers a brief review of evidence for processing differences between homographs and non-homographs in word naming and lexical decision tasks, followed by a description of the assumptions and procedures employed in construction of the database.

4.2 Homographs in psycholinguistic experiments

Homographs have been used in sentence contexts and as isolated words to investigate a variety of psycholinguistic phenomena, such as, in English, the role of meaning frequency (e.g., Simpson & Burgess, 1985), how lexical knowledge is organised (Nelson, McEvoy, & Schreiber, 2004), and inhibition of meanings in reading and word recognition (e.g., Was, 2011). However, almost all the homographs considered in English were homophonic and the cognitive processes involved in reading them may be different. As Gee and Harris (2010) note, heterophonic homographs are a small, unrepresentative group in English orthography, thus hardly any studies have investigated how they are processed.

This situation is very different from Semitic languages where heterophonic homographs predominate. In Hebrew, heterophonic homographs have been used to examine the actual processes of disambiguation including morphological decomposition (e.g., Miller, Liran-Hazan, & Vaknin, 2016) and hemispheric differences (Peleg et al., 2012). Some studies have found lexical decision in a Semitic language was faster for unvowelised homographs (which could be read as either low- or high-frequency words) than one of their vowelised versions (e.g. Bentin & Frost, 1987). This has been explained by adults being familiar with reading unvowelised text, so are slower to read vowelised text which they encounter less often. There is still uncertainty regarding which variables are responsible for differences in word naming and lexical decision latencies between unvowelised heterophonic homographs and non-homographs in Semitic languages.

4.3 Initial word selection

In the present study, an initial list of more than 2,173 candidate unvowelised letter strings was compiled from MSA textbooks (for high frequency words) and from individuals highly qualified in Arabic (for low frequency words). The MSA textbooks were widely used with adult non-native students of Arabic, thus contained many of the most commonly used words. The textbooks used were Al`Arabiyyatu Baina Yadaik (Al Fawzaan, Hussein, & Fadh, 2003) and A New Arabic Grammar of the Written Language (Haywood & Nahmad, 1962). For low frequency unvowelised letter strings, fifteen native speaking volunteers drawn from a variety of academic fields (and including the researcher) were asked to name homographic words which occur less often, but would nonetheless be comprehensible to the average native speaker, for example because they would hear or read them in the news, or use them at home when talking about work or education. The same fifteen people were asked to provide additional suggestions for high frequency unvowelised letter strings representing words that average native speakers would know from similar sources. Suggestions from volunteers were used to increase the range of vocabulary beyond beginner level and ensure a range of frequencies, in addition to ensuring that selected items were likely to represent real usage by native speakers. Items were only included when there was unanimous agreement and some 300 items were included in this way.

The classification of these words according to frequency was at this stage arbitrary and objective: arbitrary in that the thresholds were non-negotiable and adopted from Brysbaert and New (2009), objective in that they were objective frequencies as calculated by the Aralex software (Boudelaa & Marslen-Wilson, 2010), reflecting its corpus of 40 million words found in newspapers online. For the purposes of the present study, low frequencies were classified as less than 10 occurrences in a million words, while high frequencies were more than 20 occurrences in a million words, in accordance with the thresholds used by Brysbaert and New (2009).

In order to ensure that these words were classified in the minds of native speakers in terms of low or high frequency as they had been classified in Aralex, each item was presented to a panel of six adult native speakers of Arabic. The acceptance criterion for any given word was that four of the six members of the panel had to agree with the Aralex frequency classification. If this criterion was not met, the item was discarded. This process helped to ensure that high frequency words had both a high objective frequency and subjective frequency, and likewise that low frequency words had a similar measure of objective and subjective agreement. Accuracy and precision were paramount in the word selection process for two reasons: the main experiments in the present thesis would concern recognition of individual words, and other researchers using the database in the future would expect this.

Certain categories of words were excluded from the list for two reasons: some could restrict the number of vowelised forms associated with a particular unvowelised letter string, while others could distract participants from disambiguation by causing unnecessary confusion. Any word with the prefix *AL* (ال) had to be omitted due to the fact that this prefix, the meaning of which approximates to *the*, can only appear on nominal words i.e. nouns and adjectives. Thus, the sequence القدر A-L-Q-D-R could be read as *ALQaDR* (the extent) or *ALQiDR* (the pot) but not as *QaDaRa* (he was able), restricting the number of possible forms and meanings for this homograph.

Plurals were also omitted due to the fact inclusion of some but not all types of plural could have redirected attention, for example leading participants to wonder why there were so few plurals and to think about grammar instead of alternative orthographic forms. Most masculine plurals undergo a spelling change depending on the grammatical case. According to Ryding (2005) the sound (as opposed to broken) masculine plural ends with ونَ *uuna* in the nominative case and in ينَ *iina* in the accusative and genitive cases. Similarly, a spelling change occurs with the dual form, which is the 'we two' or 'you two' plural, which ends with *aani* in the nominative case and *ayni* in the accusative and genitive (Ryding, 2005). In order to maximise the opportunities for identification of homographs, in addition to avoiding potential confusion arising from points of Arabic grammar for the participants, all plurals were omitted. In terms of the effect of dual and multiple plurals on reading in Arabic, it would be more appropriate for this to be the specific focus of a separate study in the future. Upon completion of the selection procedure, the final total number of items was 1,474.

Letter strings were classified as short if they contained four letters or fewer, and as long if they contained five letters or more. These thresholds were set in accordance with the nature of Arabic morphology. Abu-Rabia and Taha (2006) state that most nouns and verbs in Arabic are constructed on a base of three consonants. Thus, a past tense verb will rarely consist of fewer than three letters. For example, كَتَبَ, pronounced KaTaBa (he wrote) consists of three letters K-T-B, with the vowel sounds achieved by the addition of diacritics. This is the shortest form of the trilateral verb that exists. In order to conjugate the present tense, a minimum of one letter will be added to those three root letters. For example, the present tense of كَتَبَ KaTaBa (he wrote) is يَكْتُبُ YaKTuBu (he writes/is writing), consisting of the four letters Y-K-T-B. Therefore, words with four letters had to be classified as short, otherwise, the vast majority of verbs in the Arabic language would have been excluded from appropriate categorisation as short in the corpus.

4.4 Name agreement

Since unvowelised letter strings can correspond to multiple vowelised forms with distinct pronunciations and meanings (heterophonic homographs), they may generate different responses in word naming. Thus, it is important to establish the normative pronunciation for each of the letter strings. Historically, researchers have employed various ways of establishing norms of name agreement to determine meaning frequency or dominance (Gee & Harris, 2010). In English, Gorfein, Viviani, and Leddo (1982) derived information about dominance by obtaining four continuous word associations to each of 107 homographs from a cohort of 100 participants (50 male and 50 female). Participants were instructed to write for each homograph the first word it made them think of, then the second third and fourth.

Gee and Harris (2010) adopted a different approach in norming 197 homophonic homographs taken from a pre-existing homograph database. First, a panel of three researchers used previously listed meanings and a dictionary to establish two main meaning categories for each word, such as ‘correct’ and ‘direction’ for ‘right’. 2,257 undergraduate participants aged 18 or over were instructed to put the word associates into as many meaning categories as applied.

In Hebrew, Peleg and Eviatar (2009) used a three-stage process to select homographs. 50 participants were instructed to circle the most frequent meaning among the alternatives provided; the meaning chosen by 65% or more of the participants was taken as the dominant meaning. Next, 50 different participants were instructed to write down the first associated word that occurred to them when reading each homograph. Where the result for 65% or more of both sets of participants were the same, the homographs were included in the experiment.

Name agreement studies have also been widely used in picture naming, a procedure which is considered to give more direct lexical access than word naming because it bypasses orthography. Two main measures of name agreement have generally been used, percentage participant agreement, and the *H*-statistic which calculates the distribution of different names given to the pictures, as described in Snodgrass and Vanderwart (1980) and so can be said to measure the extent of disagreement about a particular name. Both measures are used in the present study and are described in more detail in section 4.6.4.

4.4.1 Initial questionnaire design and piloting the procedures

The original list of 2,173 Arabic unvowelised letter strings was reduced to 1,474 as a result of initial selection followed by removal of items which could restrict alternative possibilities. Lengths were defined as number of printed letters in a string (counted by the researcher) and frequencies were checked using the Aralex database (Boudelaa & Marslen-Wilson, 2010).

Six native Arabic speakers with postgraduate level language skills piloted the initial questionnaire and name agreement procedures. They were each asked to complete two tasks. The first task asked them to consider a subset of 200 unvowelised letter strings as a letter string in Modern Standard Arabic. The second asked them to consider the same set of letter strings as representing words in the Arabic dialect spoken in the central region of Saudi Arabia.

In the first task, the six participants were instructed that letter sequences would be given to them without diacritics and that it was their task to extract as many different possibilities as they were able to for each sequence of letters by adding and adjusting the diacritics. The participants were given an example to clarify what was required in the first task. It was the sequence بيت B-Y-T presented without diacritics. They were then told that it could be read in a number of different ways and examples were given complete with diacritics. Thus, it could be read as بَيْت BaYT or بَيَّتَ BaYYaTa. It was then explained that BaYT has at least two meanings: house and a verse of poetry. So, the participants were asked to write the different forms of each sequence of letters, but were told that if they could only think of one form, that would be fine as the experimenter simply wanted to know what came to their minds. The example of قال Q-A-L was given, saying that this sequence of letters would only bring one form to the mind of the average participant: قَالَ QaALa. It was also clarified that, as in this example, each form might only have one meaning in their minds and that this too was of interest to the researcher. They were given the same example, قَالَ QaALa, which can only mean *he said*. But they were told that if they felt that one form had more than one meaning, they should write any additional meanings in the spaces provided, as a synonym or short definition or description. They were referred back to the above example of B-Y-T.

In the second task, participants were asked to identify which sequences of letters in the first task were homophones. They were then asked to think of what the words meant in their own dialect and not to be concerned if the meaning was the same as MSA. They were asked to identify whether: the word/form had the same meaning and pronunciation in their dialect as in MSA; if the word/form had the same meaning but a different pronunciation in colloquial Arabic compared to MSA; if the form (or forms) used in colloquial Arabic had the same pronunciation but a different meaning in MSA. They were also asked to identify the form (or forms) that exist in in MSA but not in colloquial Arabic and, finally, to identify the form(s) existing in colloquial Arabic but not in MSA.

Feedback from five of the six participants indicated that the questionnaire was too long and was moreover difficult to follow. They suggested excluding the section concerned with colloquial dialect and simplifying the instructions as much as possible. The relationship between colloquial Arabic and MSA is of undeniable importance (e.g. Saiegh-Haddad, 2005), but making the required judgments about words would require a subject expert. It was therefore decided that it could be dealt with in a separate study targeted at linguists and that the present study should have the section dealing with colloquial Arabic removed. The questionnaire was shortened so that it would take between 60 and 90 minutes to complete. Instructions were simplified in accordance with participant feedback.

The original intention was to achieve a list containing both heterophonic and homophonic homographs and to have participants provide information about homophonic homographs which could be analysed for use in future studies. However, it was decided to exclude homophonic homographs from the present study. The relationships between orthography, morphology and phonology in individual Arabic nouns are a matter for judgment by language experts (e.g., Benoussi, 2011). Without the presence of appropriate co-researchers who could clarify these relationships, extra items could not have added any useful dimension to the present study because it would not have been possible to determine appropriate measures for more detailed experiments.

4.5 Method

4.5.1 Participants

A total of 445 adult speakers of Arabic contributed data in the name agreement task (201 males and 244 females, mean age 29.79 years, standard deviation 9.731, range 18-60 years). All were native speakers and had normal or corrected-to-normal vision. None had any language problems.

Ethical approval was obtained from Lancaster University as well as Imam Muhammad University in Riyadh, Saudi Arabia (see Appendix C). Information sheets were distributed to all participants stating all of their rights including the right to withdraw at any time. Information was given about the purpose of the experiment, complaints procedure, expected duration of the experiment and the researcher's contact details. Each participant was reminded of their right to withdraw before they signed the consent form.

4.5.2 Name agreement materials

Normative data on vowelisation were needed to understand how native speakers read homographic letter strings, and to be able to measure and interpret the results of experiments. The task was too large for any one individual to complete, and therefore the 1,474 unvowelised letter strings were divided into 11 subsets of 98 letter strings and 4 of 99. These were presented to adult speakers of Arabic in questionnaire form and the randomisation facility on Microsoft Excel was used to generate the words contained in each questionnaire in order to reduce the likelihood of responses being influenced by the assignment or order of items. The randomisation function was programmed with six points, the first four based on the researcher's subjective judgment, to try to ensure distribution of letter strings with different relevant features in each questionnaire:

1. The word is not a homograph (it has only one possible combination of diacritics/form) and only one meaning
2. The word is a homograph (it has more than one possible combination of diacritics/form), but each form has only one meaning
3. The word is a homograph and each form has more than one meaning
4. The word is not a homograph, but it has more than one meaning
5. The word frequency (low/high)
6. The word length (short/long).

4.5.3 Distribution of materials

In order to be sure of identifying a sufficient number of participants who would form a satisfactorily representative cross-section of men and women across different age groups and educational backgrounds, the researcher enlisted the help of volunteer assistants in distributing the questionnaires. Twenty seven volunteer assistants (10 female and 17 male) were responsible for distributing the questionnaires. Each assistant was sent a set of 15 different versions of the questionnaire and was asked to distribute at least one copy of each and more if they were able to do so. Sets were coded in order to identify the volunteer and the number of the set of questionnaires, because some volunteers distributed more than one set. All the assistants participated in the study by completing one questionnaire themselves; this also ensured that they fully understood the purpose of the experiment and the instructions in the event of any issues arising with the other participants.

The assistants were instructed to give prospective participants the information sheet first, after which, if the individual agreed to participate in the study, he/she was to contact the assistant. This did not present any issues as the participants were mostly friends or relatives of the assistants, provided that they met the selection criteria. Altogether, 1005 questionnaires were distributed in 67 sets of 15 questionnaires; 554 completed questionnaires were returned, a response rate of just over 55%.

4.5.4 Name agreement procedures

Participants were asked to produce as many different vowelised forms (written with diacritics) as they knew for each unvowelised letter string (printed in standard format without diacritics) in order to establish which letter strings they considered to be homographic. They were informed that, in the questionnaire, when it came to diacritics, they should use the diacritics dhammah◌ُ fatHah ◌َ kasrah◌ِ shaddah◌ّ sukoon◌ْ. Words were to be written in MSA and participants must not refer to a dictionary or consult anybody else. They were also instructed not to write the diacritics on final letters unless they were essential to a different vowelisation and meaning of the letter string. Examples were given to clarify these points. They were also advised to be wary of confusing the letter taa marbooTah ة with the letter haa ه as the two letters have the same pronunciation at the end of the word if the word is read in isolation or occurs before a pause, for example at the end of a sentence.

Personal information was requested, including gender, age, and level of education, occupation, and region of Saudi Arabia.

4.5.5 Data organisation

For the name agreement, of the total of 1005 questionnaires, 554 were returned, of which 109 questionnaires (just under 20%) were disregarded for one or more of the following reasons:

1. Diacritics had been added at random, making no sense, as if the participant was not interested in the integrity of his or her responses.
2. The diacritics followed a pattern as though the participant thought that they were simply required to put a different set of diacritics on every occasion whether this resulted in a real word or not.
3. The participant had not understood what was required and wrote either a synonym or an antonym of the given word.
4. The participant was outside the desired age bracket of 18 to 60 years.
5. The words had been written without diacritics.

The remaining data were input to a worksheet. The first column contained the unvowelised letter string (the target item). The second contained the participant response for that letter string. If more than one response was given to a letter string, each response was listed on a separate row. The third column contained the correct fully diacritised response as listed in the dictionary for each of the responses given by the participant. The fourth column contained the modal response for each unvowelised letter string.

The next two columns were used to record the researcher's evaluation of the participant's response and the type of error (if errors were made) and the experimenter cross-checked with the dictionary before entering the correct or corrected form. The researcher’s evaluation of the participant’s entry was based on four possibilities: correct, shortage, redundancy and wrong. ‘Correct’ indicated there was absolutely no error or shortage, while ‘shortage’ indicated that something was incomplete, typically meaning that not all of the required diacritics are present. A word with a shortage could be acceptable or unacceptable; in order to be classified as accepted, the diacritics present excluded all other possibilities. ‘Redundant’ was used for superfluous details, such as diacritics when not needed at the end of the word (as this is determined by grammatical rules which in turn are determined by context). Redundancy was also classified as acceptable or unacceptable depending on whether all other possible forms of the word were excluded. ‘Wrong’ indicated that the answer given was a non-word. A separate column was used to record the number of fully vowelised forms returned by the participant for that word and another column recorded how many different diacritised forms the participant had identified for each unvowelised letter string.

Separate columns were also used to record the number of responses and percentages of responses agreeing on a particular form, followed by a column showing the total number of responses. Another column was used to record the number of different meanings given for a specific vowelised form, where more than one meaning was given (homophonic homograph).

Three columns were used to record the codes of the participant, contact (assistant) and questionnaire number. Demographic information was recorded in 5 columns for gender, age, education, occupation and region of Saudi Arabia. Further columns were used to record situations in which participants used MSA: work, home, with friends, and social media.

Finally, two columns were used to record the length in letters of the unvowelised letter string and its frequency as extracted from the Aralex database (Boudelaa & Marslen-Wilson, 2010).

The fact that many participants in the name agreement experiment did not fully vowelise all the letter strings given may indicate that they did not know those forms of the words, or perhaps that the instructions were not fully understood. Moreover, many mistakes were observed in spellings; participants frequently confused ta marbootah (ة) with ha (ه), likely to be because they have the same pronunciation when the word is isolated or at the end of a sentence. In addition, there was some confusion between the ya (ي) and alif maqsoorah (ى), assumed to be due to the similarity in appearance.

4.6 Analysis of results

This section presents appropriate statistical procedures to analyse participant characteristics (gender, age and education) which could affect the number of vowelised forms produced by each participant, followed by item characteristics (frequency and length of letter string in letters) which could also affect name agreement. T-tests were used for comparison of categorical (binary) to continuous measures, chi-square ($\chi 2$) for comparison of two binary variables, and point-biserial correlation for measurement of the strength of association between a binary and a continuous variable. The phi (φ) coefficient measure of association, which is interpreted in a similar way to the correlation coefficient, was used for the two binary variables (gender and education).

Name agreement was calculated using three measures: percentage name agreement, type (the number of vowelised alternatives identified for each letter string) and the *H*-statistic. The *H*-statistic takes into account the distribution of responses across all correctly diacritised forms and participants. Details of how name agreement and the *H*-statistic were calculated are given in section 4.6.4. Multiple linear regression was employed with name agreement and the *H*-statistic to discover which variables constituted a model that fitted the data more closely.

4.6.1 Participants

Participant characteristics, in terms of gender, age and education, were analysed using t-test and chi-square analyses to ensure that any bias resulting from demographic data could be identified and addressed. T-tests were used where one variable was continuous and the other binary, and chi-square analysis where both variables were binary. The mean count of all correctly diacritised alternatives for each of the 1,474 letter strings was calculated for each participant. The total sample size for the name agreement survey consisted of 445 participants (201 males and 244 females, 45.3% and 54.7% respectively).

In terms of the level of education, participants were allocated to one of two groups: graduate and non-graduate. The distinction between the levels was made on the basis of the probable extent and depth of the individual's exposure to MSA. 'Graduate' included all those who had obtained a bachelor's degree or above at any institution, while 'non-graduate' included everyone who had completed their education before obtaining a bachelor's degree. There were 253 graduate participants (56.9% of the sample) and 192 non-graduates (43.1%).

Information regarding the situations in which participants typically used MSA as distinct from colloquial Arabic was collected in four broad categories: work, home, friends and social media. Participants were asked if they employed MSA in each of these situations. Interestingly, many more participants used MSA in the context of social media (303) than in a work situation (193). This may reflect the fact that many Saudi Arabian women are at home rather than at work, thus their use of MSA relates to social media. Very few participants used MSA at home or with friends (10 and 16 respectively).

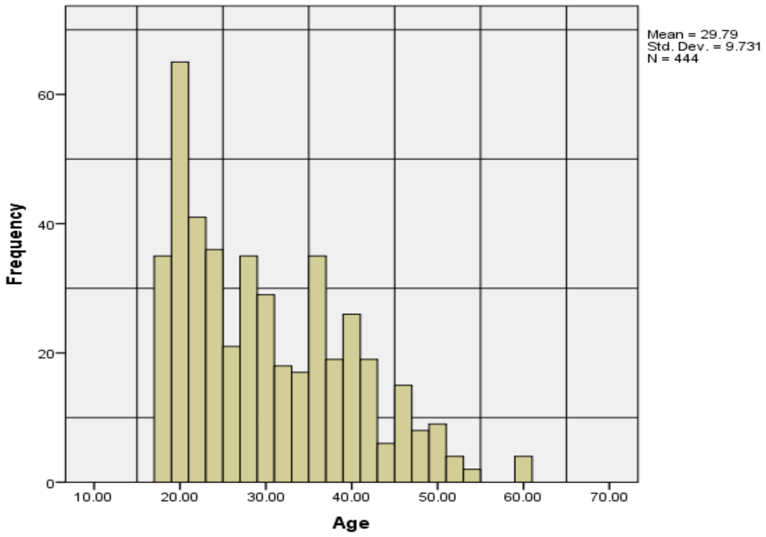


Figure 4 Distribution of name agreement study participants by age

Figure 4 shows that there were more participants in their 20s than any other age group. This probably reflects the fact that most of the volunteer assistants who distributed the questionnaires were themselves in that age group.

4.6.2 Relationships between variables

The results of t-tests and chi-square analyses indicate how the participant attributes of age, gender and education relate to each other, and to the outcome variable of mean count (of correctly diacritised alternative forms of the letter strings per participant).

There was a significant difference in the ages of females ($M = 27.9$, $SD = 8.6$) and males ($M = 32.1$, $SD = 10.5$); $t(442) = -4.64$, $p = < .001$.

No association was found between gender and level of education χ^2 (1, *N* = 445) = .825, p = .364, but there was an association between age and education. There was a significant difference in the age of graduates (M = 33.45, SD = 8.44) and non-graduates (M = 24.94, SD = 9.20); t(442) = -10.11, p = < .001.

There was no significant different in the mean count of females (M = 1.47, SD = .26) and males (M = 1.49, SD = .34); t(443) = -.503, p = .62.

There was a significant effect for education t(443) = -4.92, p = < .001 with graduates (M = 1.5, SD = .32) identifying more vowelisations per item than non-graduates (M = 1.4, SD = .25).

Table 5 Pearson's r bivariate and point bi-serial correlations between key demographic variables (N = 445 except for age where N = 444)

		Age	Gender	Education group
Mean count	r	.237**	.024	.227**
	p	< .001	.615	< .001
Age	r		.216**	.433**
	p		< .001	< .001

Notes: 1. mean count = mean count per participant of correctly diacritised responses
2. education group = graduate or non-graduate

There was no association between gender and education, Φ = .043, p = .364.

It can be seen in Table 5 that there was a weak positive correlation between mean count and age (r = .237, p < .001); a higher number of correctly diacritised forms produced by each participant was associated with higher age. A weak positive correlation between mean count and education group (r_{pb} = .227, p < .001) indicated that a higher number of correctly diacritised forms produced by each participant was associated with graduate level education. There was no association between mean count and gender (r_{pb} = .024, p = .615). There was also a moderate positive correlation between age and level of education (r_{pb} = .433, p < .001); graduate level education was associated with higher age. Finally, there was a weak correlation between gender and age (r_{pb} = .216, p < .001), men being associated with higher age. The correlations between these variables support the view that participants had acquired a larger MSA lexicon over their lifetime, in addition to illustrating the expected link between education and greater exposure to MSA.

4.6.3 Regression

Age was converted to a standardised z score before running the regression. A significant regression equation was found (F(3, 440) = 12.082, p < .001), with an R^2 of .076. Results of the regression analysis indicated that two of the three variables explained most of the variance, age and level of education. Effects estimates can be seen in Table 6.

Table 6 Estimates of effects

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	1.433	.026		55.004	< .001
zscore(Age)	.052	.016	.174	3.339	< .001
Gender	-.011	.028	-.019	-.403	.687
Education	.093	.031	.154	3.031	.003

The estimated coefficient implies that the mean count of correctly diacritised forms changes by .093 for unit increase in education from non-graduate to graduate. For unit increase in education from non-graduate to graduate level, the number of correctly diacritised forms increased by .093 ($\beta = .154$, $SE = .031$, $t = 3.031$, $p = .003$). For every unit increase in age, the mean count of correctly diacritised alternative forms increased by .052 ($\beta = .174$, $SE = .016$, $t = 3.339$, $p < .001$).

4.6.4 Item analysis

To conduct the analysis of the target words, survey results were input to a master spreadsheet in Excel to allow for checking accuracy of data and calculating name agreement and the *H* statistic, after which data were input into IBM SPSS statistical software (version 22) for analysis. The following variables were considered: length, frequency, name agreement, word type and the *H*-statistic.

Letter string length ranged from 2 to 8 characters ($M = 4.15$, $SD = 1.03$) and frequency from .01 to 2854.84 per million ($M = 63.01$, $SD = 161.94$). Figure 5 shows the distribution of letter string lengths and Figure 6 shows the distribution of frequencis.

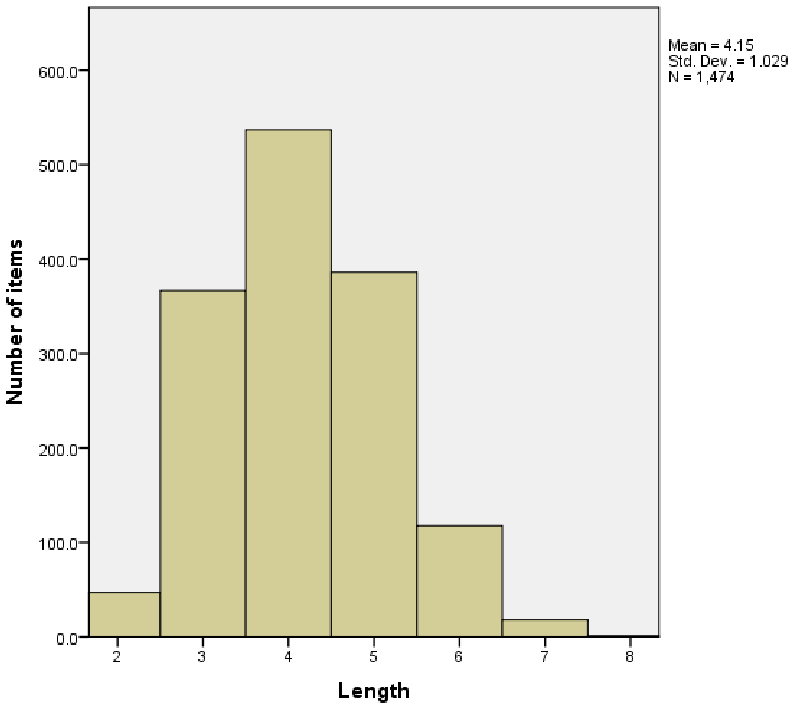


Figure 5 Distribution of letter string lengths

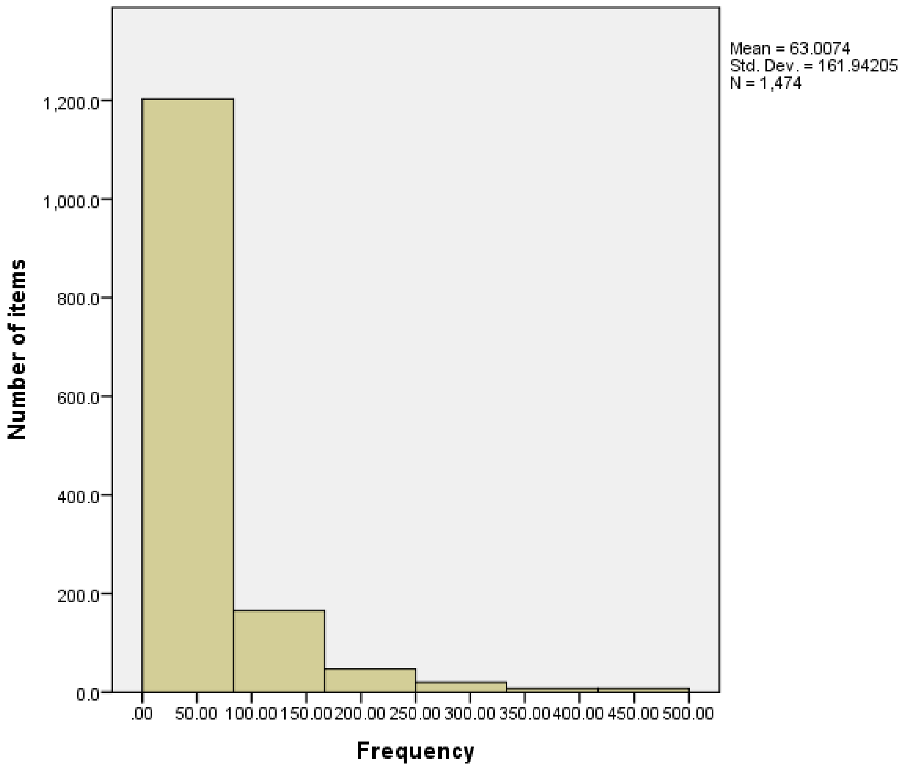


Figure 6 Distribution of frequencies

Calculations of percentage name agreement and the *H* statistic were as follows. Percentage name agreement was calculated on the basis of the modal response (dominant meaning) to each letter string as follows. For each target item, the total number of responses for all the possible correctly diacritised responses was calculated, and the total number of each of the possible alternatives calculated as a percentage of the total. The percentage name agreement was the highest proportion of total correct responses (modal response) to the target item. The highest percentage ranged from 21% to 97% agreement for the homographic items, while non-homographic items exhibited 100% name agreement. Levels of agreement were input as figures ranging from .21 to 1, where 1 indicated 100% agreement among participant responses. The mean value for the name agreement variable was .76 (76%) with a standard deviation of .236. Figure 7 shows that name agreement was 100% for 587 of the 1,474 items.

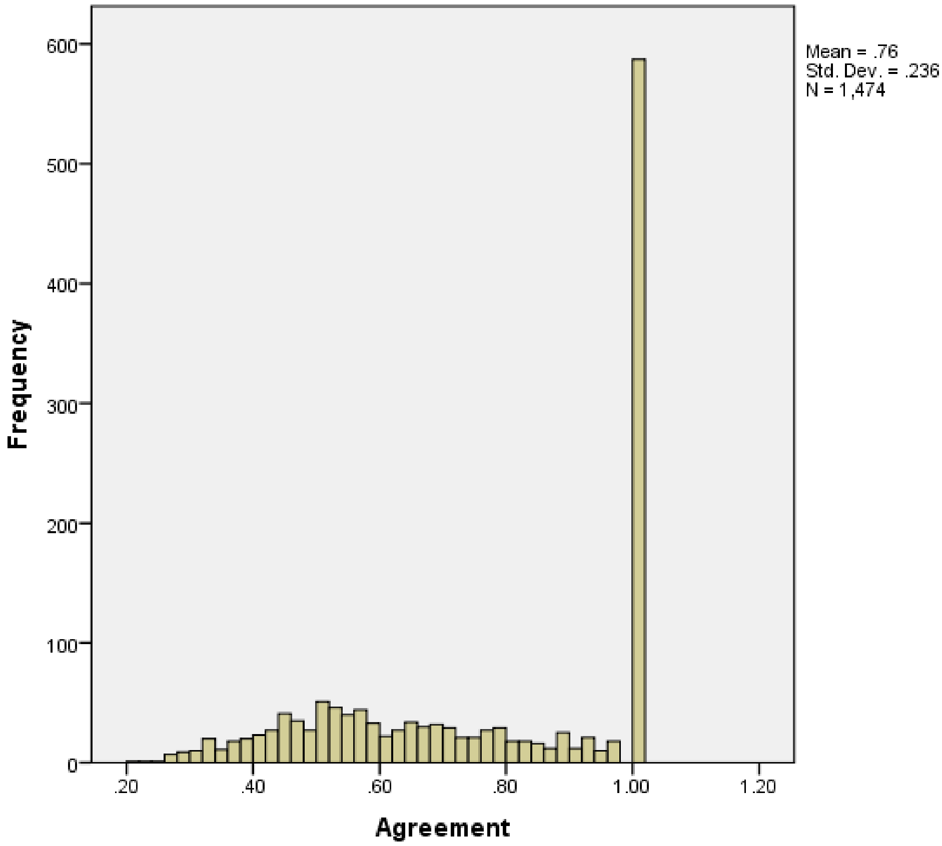


Figure 7 Distribution of percentage name agreement

In order to further inspect the homographic items, items showing 100% agreement were removed, and a new sample size (N = 887) analysed. This generated a mean name agreement value of .61 with a standard deviation of .176. The resulting distribution of percentage name agreements is presented in Figure 8, showing that the level of agreement differed according to the number of correctly diacritised forms of the homographic letter string.

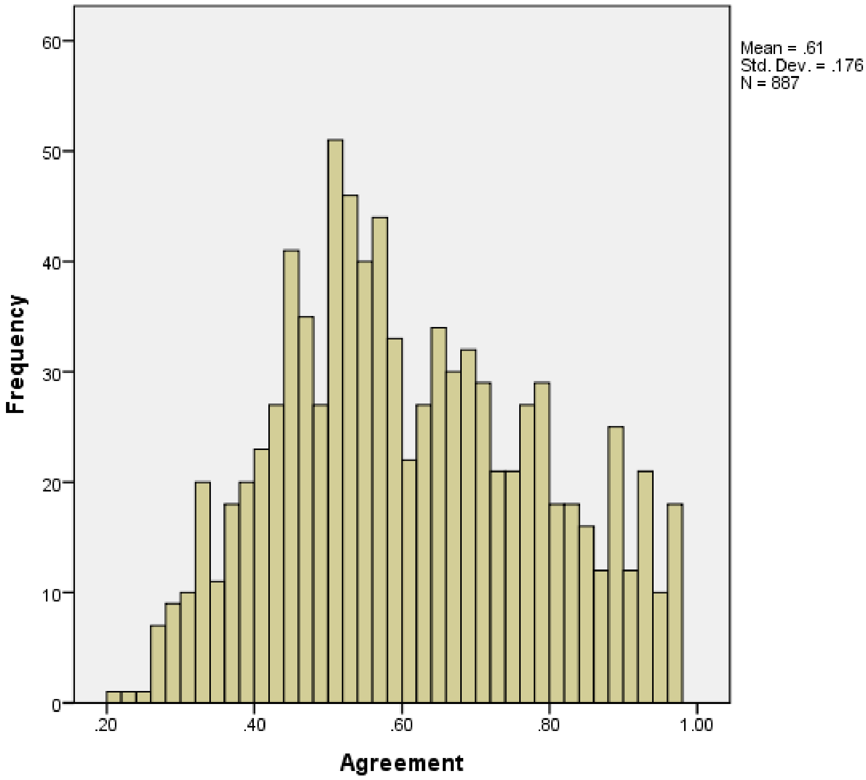


Figure 8 Distribution of name agreement for homographic items

The information statistic (*H*), (also referred to as the 'U statistic'), measures name agreement differently, by taking into consideration the proportion of subjects producing each of the correctly diacritised responses (types), not just the modal response. For each target item, the number of correctly diacritised responses was counted for every alternative type and the number of responses for each type calculated as a percentage of the total correct responses to the target item. The number of alternative types ranged from 1 to 11, depending on the target item.

The information statistic *H* was then computed for each target word by the formula:

$$H = \sum_{i=1}^{k} p_i \log_2(1/p_i)$$

where *k* refers to the number of different uses of correct forms for each type (correctly diacritised alternative), P_i is the proportion of participants who produce that form and Log_2 is the logarithm base 2 of $(1/p_i)$.

The *H* statistic is a continuous variable which can take values from 0 upwards. A word that has a high significant difference between the target item and the word produced, would have an *H* value of 1.00. For example, a word with two forms and meanings identified by 50% of participants in each case would have an *H* value of 1.00 (Snodgrass & Vanderwart, 1980). A higher *H* value corresponds to a lower level of name agreement and, by implication, a lower proportion of participants who all gave the same word form as their first alternative. In this way, the *H* value captures more information about how the target word is distributed across subjects: two homographs could have the same percentage score for name agreement but different *H* values if the number of alternative forms (types) differs. For two homographs with the same percentage score, the homograph with more alternatives would result in a higher value of *H*.

Five hundred and eighty seven (587) values of *H* were zero, due to a single form of diacritisation being used by all participants for these target items, and corresponding to non-homographic words. Thus, the distribution was not a normal distribution when all target words were included. After removal of words where $H=0$, the resulting distribution of *H* values was as shown in Figure 9.

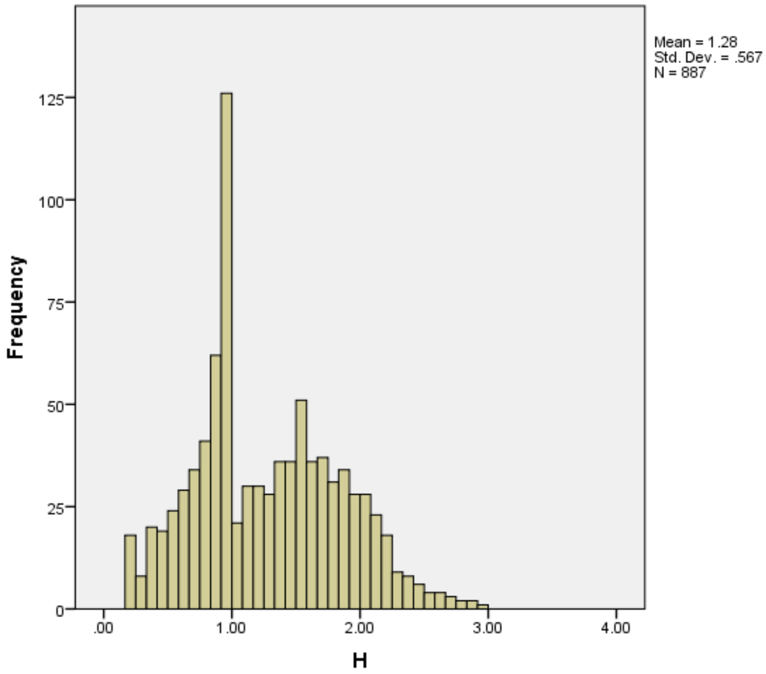


Figure 9 Distribution of *H* statistics for homographs

Actual length in letters of each unvowelised target letter string was used as a continuous variable. A summary of target item characteristics is shown in Table 7.

Table 7 Mean, standard deviation, minimum and maximum, and range of variables (N = 1,474)

Variable	Mean	SD	Min.	Max.	Range
Agreement	.78	.23	.21	1.00	.79
H	.72	.75	.001	2.94	2.94
Type	2.40	1.73	1.00	11.00	10
Frequency	64.13	158.97	.01	2854.84	2854.83
Length	4.17	1.03	2	8	6

It can be seen in Table 7 that the mean type count (number of correctly diacritised alternatives) for all items was 2.4. After removing the non-homographs from the calculation, there remained 3,079 different correctly diacritised forms for the 887 homographic items, giving a mean type count (number of correctly diacritised alternatives) of 3.47.

4.6.5 Correlations

Correlations were used to test for associations between the predictor variables *H*, length, type and frequency variables and the relative strengths and directions of any such associations. Correlations between the predictor variables of length and frequency were used to examine how they related to each other and check for collinearity, while correlations between predictor and outcome variables (alternate measures of name agreement) were also examined.

Table 8 Summary of Pearson's r bivariate correlations of item variables (N = 1,474)

		H	Frequency	Length	Type
Agreement	r	-.969**	-.075**	.495**	-.838**
	p	< .001	.004	< .001	< .001
H	r		.76**	-.517**	.929**
	p		.004	< .001	< .001
Frequency	r			-.148**	.0720**
	p			< .001	.006
Length	r				-.504**
	p				< .001

Notes: 1. Agreement = percentage name agreement 2. *H* = *H*-statistic 3. Frequency = frequency of unvowelised letter string from Aralex (Boudelaa & Marslen-Wilson, 2010) 4. Length = length in letters of unvowelised letter string 5. Type = number of different correctly diacritised forms of letter string

There was a strong negative correlation between agreement and *H* (r = -.969, p < .001), with *H* approaching 0 as agreement approached 100%. An increase in percentage agreement was associated with a decrease in the number of alternative forms of diacritisation and number of participants agreeing on them. A weak negative correlation between agreement and frequency ($r = .075, p < .001$) indicated that an increase in frequency was associated with a decrease in percentage name agreement, while a moderate positive correlation between agreement and length ($r = .495, p < .001$) indicated that longer letter strings elicited greater name agreement. A strong negative correlation between agreement and type ($r = -.838, p < .001$) indicated that agreement decreased as the number of correctly diacritised alternatives for a homographic letter string increased.

For the *H* statistic, there was a weak positive correlation with frequency ($r = .076, p< .001$), a decrease in frequency associated with a decrease in the extent of name agreement across the number of correctly diacritised alternatives; higher frequency words were associated with a greater number of possible alternatives. A moderate negative correlation between the *H* statistic and length ($r = -.517, p < .001$), indicated that shorter words were associated with more possible alternatives. There was a strong positive correlation between *H* and type ($r = .929, p < .001$), with an increasing value of *H* increasing associated with a higher number of possible alternative diacritised forms of the letter string.

A very weak negative correlation between frequency and length ($r = -.148, p< .001$) showed an increase in length associated with a decrease in frequency, while a very weak positive correlation between frequency and type ($r = .072, p = .006$) indicated that an increase in frequency was associated with an increase in the number of correctly diacritised forms of the letter string. A moderate negative correlation between length and type ($r = .504, p< .001$) indicated that an increase in length was associated with a decrease in the number of correctly diacritised forms of the letter string.

4.6.6 Regression

Multiple linear regression was carried out in order to predict the value of a dependent variable based on the value of two or more other, independent variables. Multiple regression allows identification of both the overall fit, and the variance to be explained, between data and model, and identifies the relative contribution of each of the predictors to the total variance. All variables were standardised as z scores before a range of regression models was run, based on the earlier elimination of variables which made no contribution to explaining the variances.

Two different regression models were run: (1) Name agreement as dependent variable, with zscore(Length) and zscore(Frequency) as predictor variables; (2) *H* as dependent variable, with zscore(Length) and zscore(Frequency) as predictor variables.

4.6.6.1 Regression model 1

A multiple regression analysis was calculated to predict *percentage name agreement* based on target item *length* and *frequency.* A significant regression equation was found (F(3, 1471) = 238.414, p = < .001), with an R^2 of .245. Output from the first regression is shown in Tables 10 and 11.

Coefficients of effects of variables (Table 9) show that, within this model, item length contributed to explanation of the variance in the mean of name agreement, and the relationship was positive. The estimated coefficient implies that agreement changes by .118 for unit increase in standardized length: longer unvowelised letter strings tend to elicit vowelised forms with higher name agreement. No change in name agreement is implied by a change in frequency.

Table 9 Survey regression model 1: Coefficients of variables

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.763	.005		142.841	<.001
	zscore(Frequency)	<.001	.005	-.002	-.090	.928
	zscore(Length)	.117	.005	.494	21.582	<.001

4.6.6.2 Regression model 2

A second regression analysis was carried out to predict *H* based on target item *length*, and *frequency* in order to determine which of the two measures (name agreement and *H*) would give a better fit between model and data. A significant regression equation was found (F(3,1470) = 178.795, p < .001), with an R^2 of .267.

Coefficient estimates from the second regression is shown in Table 10.

Table 10 Survey regression model 2: Coefficients of variables

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
2	(Constant)	.771	.017		45.080	<.001
	zscore(Frequency)	-.001	.017	-.001	-.030	.976
	zscore(Length)	-.396	.017	-.517	-22.891	<.001

Zscore(Length) significantly predicted percentage name agreement, β = .396, t = 22.891, $p < .001$. The negative relationship means that for unit increase in length, *H* is expected to decrease by .4. Zscore(Frequency) was not significant, β = < .001, t = -.030, p = .976.

The adjusted R square figure indicates that this second model accounts for almost 27% of the variance around the mean of the *H* statistic which can be explained by this model. The value of R^2 is slightly higher than in model 1, suggesting that the same predictors as in model 1 give a better prediction for *H* than for agreement. However, the higher value of the standard error of the estimate suggests that there are wider variations in the values that can be predicted from this model.

4.7 Discussion

Analysis of name agreement may provide an insight into the operation of concepts of lexical selection and is therefore potentially useful to understanding how Arabic readers process heterophonic homographs. The present study aimed to collect data pertaining to the extent to which unvowelised letter strings were considered by normal skilled Arabic readers to be homographic. Since this had not been previously studied, the aim was to collect at least 2,000 letter strings which adult normal readers would be expected to know. After elimination of letter strings which could not have been understood as homographic, such as plurals, and nouns with attached determiners, prepositions and conjunctions, a set of 1,474 unvowelised letter strings remained. This set was used to elicit the possible alternatives for each string.

Of the target items, 887 were identified as homographs by the 445 participants. The remaining 587 items had name agreement of 100%. The homographic items had a mean of 3.47 correctly diacritised alternative forms, with a range of 2 to 11 alternatives. Participants would therefore have representations of these alternative forms available in their mental lexicon when reading unvowelised text. By asking them to think of as many alternative forms as possible, it has also been established that they had a good knowledge of orthography, patterns of short vowels and how those patterns related to roots (morphology). Participants identified 1,064 letter strings as nouns and 410 as verbs; the relatively lower number of verbs may have been due to the avoidance of grammatical information indicating person and tense in the selection of stimuli.

Multiple regression analyses confirmed that longer words were associated with a higher level of name agreement and elicited fewer alternatives. This may be due to the fact they tend to have lower frequencies and therefore potentially fewer orthographic or other neighbours or competitors for lexical retrieval, although the correlation between frequency and length was weak.

Very few studies have investigated heterophonic homographs in unvowelised print in a Semitic language. Certainly, the present study offers the largest such database of heterophonic homographs in either Hebrew or Arabic. In previous studies, various methods have been used to select homographs. Bentin and Frost (1987) identified 16 homographs from a starting list of 50, each of which could represent one of two nouns which had different word frequencies. The list was reduced after 50 participants rated the frequency for each alternative in every pair, using a rating scale from 1 to 5 (least frequent to most frequent). The means of the estimated frequencies were used as a measure of agreement among paricipants to decide which homographic pairs to include in the final list.

Peleg and Eviatar (2009) used 56 heterophonic and 56 homographic homographs, all of which were nouns. The dominant or correct form was determined on the basis that 65% of the participants identified the same most frequent meaning. It appears that participants were given the choice of only two meanings and it is not known how the meanings were obtained.

Creation of a database of 2,400 fully pointed derived Hebrew nouns which shared a root, but were unrelated in meaning, with other nouns followed a similar process (Seroussi, 2011). Consultation and comparison of entries from dictionaries was followed by assessment of whether they met the semantic transparency criterion. Seroussi was assisted in this by two researchers with recognised knowledge of the Hebrew language. The assessment of subjective familiarity and subjective frequency followed a similar process to the name agreement in the present study, and concreteness was rated for a subset of some 400 nouns by a further 30 participants.

In the present study, the negative correlations between the *H* value and agreement, and *H* value and type, confirm the earlier results that high name agreement corresponds to a low *H* value and that homographs are associated with lower levels of name agreement. Correlations showed that an increase in length was associated with a decrease in the number of types and an increase in name agreement. Where individual target words have a high *H* value in experiments, it is likely that name agreement will be weak, and that this may be associated with longer latencies in word naming and lexical decision tasks, in comparison with individual words which have a low *H* value.

Participants gave alternative meanings to just over 60% of the unvowelised letter strings which implies that the homograph phenomenon may indeed be very common in Arabic, although not as high as this because the selection of letter strings deliberately avoided providing syntactical clues which would have made only one interpretation admissible. Some scholars have argued that reading without short vowels and context is impossible in Arabic (e.g., Abu-Rabia, 1997b). Others have asserted that adults read by directly recognising the unvowelised spelling as a whole word, (e.g., Ibrahim, 2013; Taouk & Coltheart, 2004), as is also seen in Hebrew (Frost et al., 1987). This claim is based on isolated unvowelised pseudoword letter strings taking longer to pronounce than isolated unvowelised real word letter strings (Ibrahim, 2013), and assumes a direct connection between unvowelised letter strings and vowelised whole words in the mental lexicon.

The results of the name agreement exercise in the present study indicate that readers recognised different vowelised possibilities from a single unvowelised letter string, which means that they are able to search these items in their mental lexicon or create them from their existing knowledge of how words are structured in Arabic. Some participants identified more possibilities for a letter string than others, which suggests they had larger vocabularies. The possible implication for the organisation of the mental lexicon in Arabic is that there are mental lexicons for both unvowelised letter strings and whole vowelised words. In addition, vowelised spellings may be accessed through identification of morphemes in the unvowelised letter string. This could be explained by the Psycholinguistic Grain Size theory (Ziegler & Goswami, 2005), if the principle of using larger grain sizes to read orthographies with less consistency between orthography and phonology is extended to using larger grain sizes because they are more efficient for reading. Printed Arabic text is consistent in its representations of words, but not at the level of grapheme to phoneme correspondence. In addition to recognising morphemes (Boudelaa & Marslen-

Wilson, 2015), Arabic readers may recognise whole words directly from letter strings.

An alternative explanation lies in connectionist theories which assume that patterns such as morphemes and unvowelised letter strings are created by repetition of other attributes of words which alter the weights on the connections over time. The word naming experiment offers an opportunity to further investigate the cognitive processes involved.

Chapter 5 A STANDARDIZED SET OF 1,474 ARABIC LETTER STRINGS: AGE OF ACQUISITION (AOA) RATINGS

5.1 Introduction

This chapter justifies the AoA study and explains how it was conducted. Procedures for this study are described, after which statistical analysis of the data is explained, and predictive variable models presented. The chapter concludes with a discussion of the findings.

Much of the literature regarding the use of MSA has emphasized the fact that it is not learned until a child begins school and that many of the common words in the colloquial dialect they use do not exist in MSA (e.g., Abu-Rabia, 2000; Ayari, 1996; Bentin & Ibrahim, 1996). In addition to the variables of word length, frequency and name agreement, AoA was therefore considered important to include, particularly as neighbourhood density measures, which are considered to affect disambiguation of homographs, were not obtainable within the scope of this research, as explained in section 3.2.3.

AoA has been shown in English to account for 4% of the variance in lexical decision reaction times which was not already accounted for by word length (measured in syllables as well as letters), frequency (log word frequency), and OLD20 (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Moreover, there is evidence for interaction between the effects of AoA and word frequency, for example in English (Kuperman et al., 2012) and in Dutch (Brysbaert, Lange, & Wijnendaele, 2000). There is also evidence for a separate effect of AoA in some languages, for instance in reading aloud in Spanish (Wilson, Cuetos, Davies, & Burani, 2013) and in picture naming in Arabic (Khwaileh, Body, & Herbert, 2014).

Some AoA researchers (e.g., Morrison & Ellis, 1995) have suggested that findings from earlier research which claimed to show large effects of word frequency may in fact demonstrate the combined effect of frequency and AoA because the latter was not controlled for. Their assertion that correlation between AoA and word frequency needs to be taken into account is one of the main reasons for investigating the effect of AoA in the present study and for collecting data on the AoA in this survey.

5.2 Method

5.2.1 Participants

Participants were recruited on-line and in total 92 individuals contributed data. Three groups were recruited. Group 1 consisted of 16 males and 17 females, mean age 31.15 (SD =8.65), 28 graduate and 5 non-graduate. Group 2 consisted of 17 males and 13 females, mean age 33.43 (SD =7.57), 26 graduate and 4 non-graduate. Group 3 consisted of 12 males and 14 females, mean age 30.96 (SD = 11.81), 20 graduate and 6 non-graduate.

Ethical approval was obtained from Lancaster University as well as Imam Muhammad University in Riyadh, Saudi Arabia (see Appendix C).

5.2.2 Materials

The first point to make is that participants were asked to rate the vowelised forms of the 1,474 letter strings on which there was highest overall name agreement when all identified correctly diacritised forms had been taken into account. This was essential to ensure that all participants were giving estimates related to the same form of the unvowelised letter string; otherwise, estimates would have been meaningless. Participants were sent the list of 1,474 letter strings which they were asked to rate.

5.2.3 Distribution of materials

Questionnaires were distributed to three groups of individuals over a period of two months. Forms which could be completed electronically were created using Googleform and the link was sent by e-mail to 30 individuals who were asked to distribute them further. The individuals in the three groups were asked to complete electronic forms for different sets of fully vowelised words: 500 words for each of two groups and 474 for the third group.

5.3 Procedures

The 1,474 word items were fully vowelised and the dominant meaning (modal response to be used in the word naming experiment as the correct answer) was used in place of each homograph. Allocation of word items to questionnaires was achieved with the randomisation function on Excel, as for name agreement.

Participants were asked to write their best estimate of how old they were (their age in years) when they learned each word, following the same instructions as Kuperman et al. (2012). It was made clear that 'learned' referred to the age when they first understood the word, even if they were not using the word themselves at that age. The present study used actual ages rather than a 7-point rating scale as used in a number of studies, for example Portuguese nouns (Marques, Fonseca, Morais, & Pinto, 2007). The 7-point rating scale would probably have introduced an extra step for some participants who would tend to think about where they were and how old they were when they learned it before locating the correct box. Moreover, it could have encouraged respondents to guess and tick a box with less careful thought.

5.4 Data organisation

For the AoA ratings, a total of 92 questionnaires were returned, 34 from the first group, 32 from the second group, and 26 from the final group. One questionnaire was disregarded in group 1, and 2 in group 2, due to the participants having misread the instructions. In each group of responses, less than 1.0% of the remaining items were disregarded. In accordance with Kuperman et al. (2011), items indicating an acquisition age of over 25 were taken out in order to remove outliers. Data were input to an Excel spreadsheet and the mean AoA was calculated for each item.

5.5 Analysis of results

Mean AoA was calculated by participant and by item. A variety of tests was employed to measure participant characteristics (gender, age and education) and detect any effect on mean AoA ratings per participant. T-tests were used where comparison was between a continuous and a categorical variable, and chi-square tests where two categorical (binary) variables were compared. Correlations included point biserial correlations for determining the strength of association between a continuous and a categorical variable, with the phi coefficient used to test for association between two binary variables. Correlation analyses were used to analyse relationships between item characteristics (frequency, length of letter string in letters, *H* and name agreement as measures of vowelisation diversity, mean AoA per item) and identify possible occurrences of collinearity. Two multiple regressions were then run, the first with AoA as the dependent variable, to examine item effects that influence by-items variation in AoA, and the second with *H*, to examine the potential influence of AoA, among other item properties, on name agreement.

5.5.1 Participants

Participant characteristics were analysed using t-test and chi-square tests in relation to gender, age and education, in order to ensure that any bias resulting from demographic data could be identified and addressed. The mean AoA for each of the 1,474 agreed vowelised letter strings was calculated for each participant and used as the dependent variable. The total sample size for the AoA ratings study consisted of 89 participants (47 males and 42 females, 51% and 49% respectively). Before presenting the correlations and regression, it useful to show the mean age and mean AoA ratings of the three groups of participants to indicate how similar they were (Table 11).

Table 11 Mean age and AoA for each group

	Mean age in years	Mean AoA in years
Group 1	31.15	10.65
Group 2	33.43	10.70
Group 3	30.96	10.36

The overall range was 9 years, which indicates that most of the MSA words in the list were learned during the years of education before entry to university.

Distributions of the AoA ratings before removal of outliers (AoA over 25 years) for each group are shown in Figures 10 to 12. Frequencies represent the total number of responses to the individual target words.

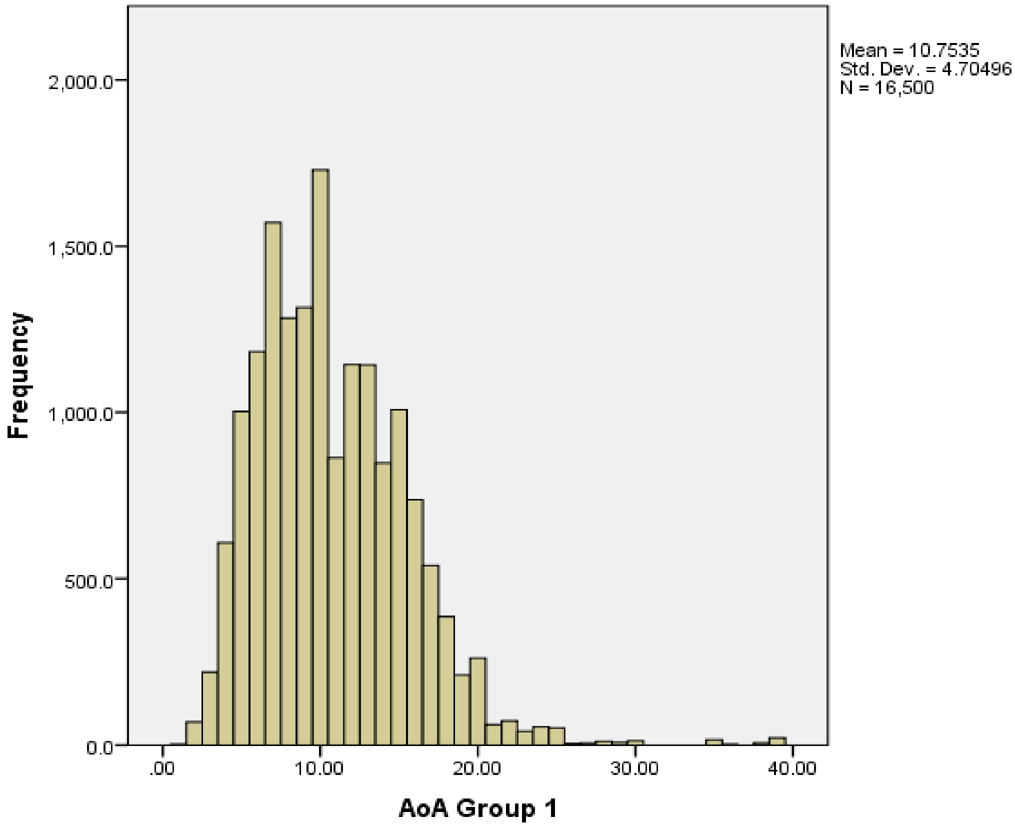


Figure 10 Distribution of mean AoA: group 1

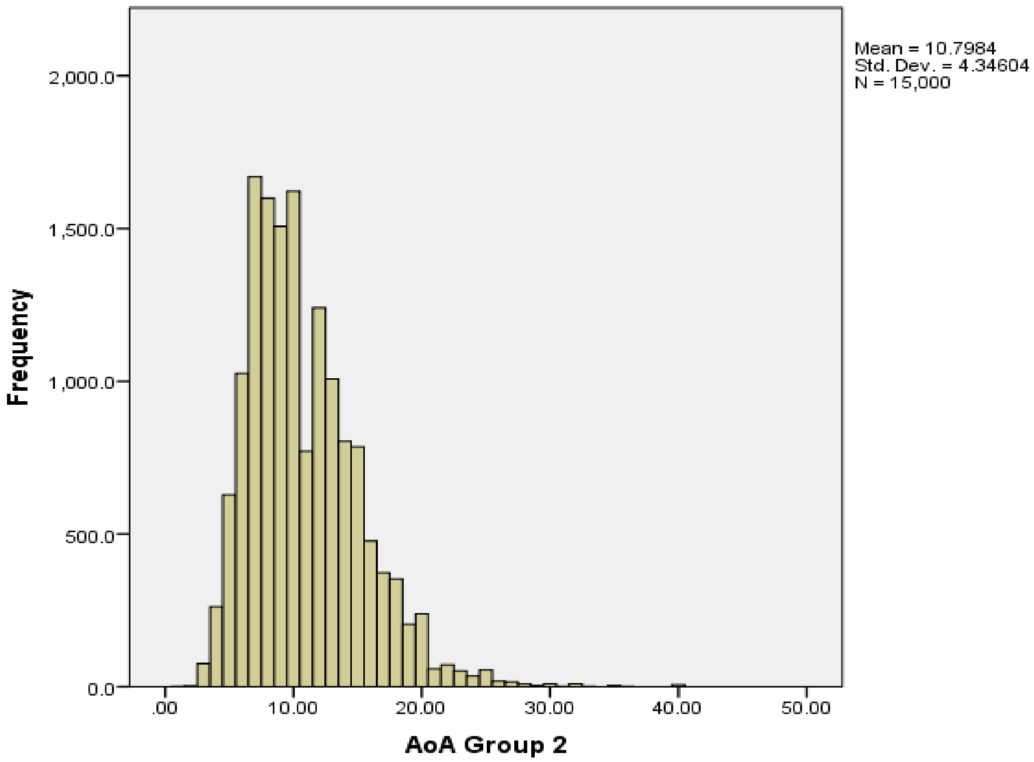


Figure 11 Distribution of mean AoA: group 2

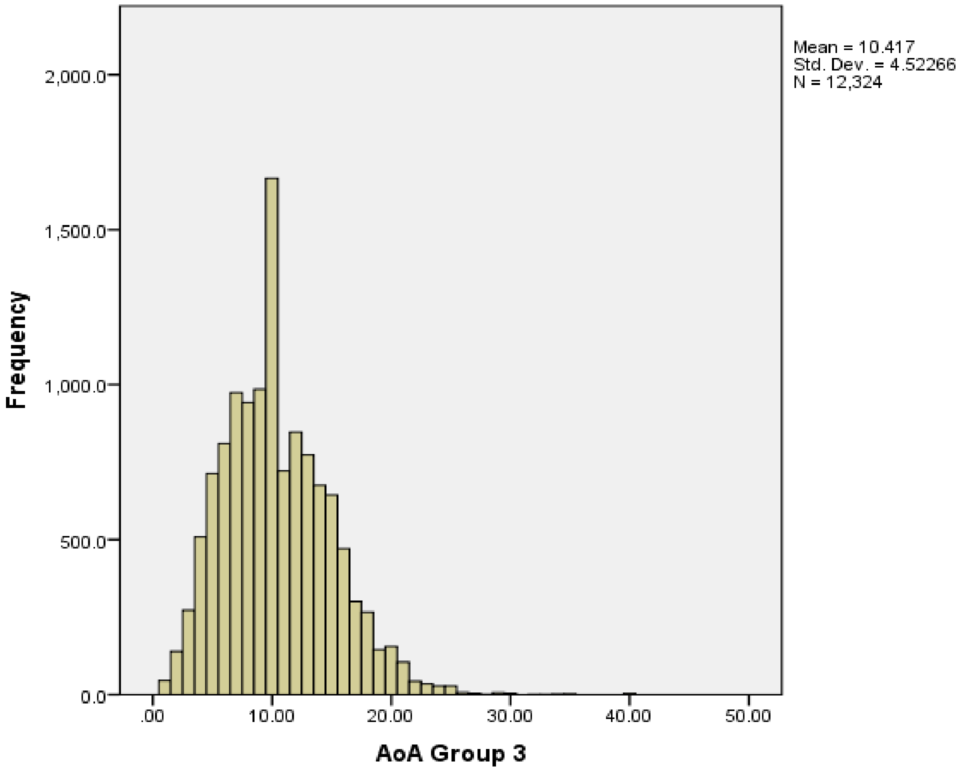


Figure 12 Distribution of mean AoA: group 3

5.5.2 Relationships between variables

The results of t-tests and chi-square indicate how the participant attributes of age, gender and education relate to each other, and to the outcome variable of mean AoA. There was a significant difference in the ages of females (M = 28.7, SD = 7.75) and males (M = 34.9, SD = 19.78); t(87) = -3.33, p = .001. No association was found between gender and level of education $\chi 2$ (1, N = 89) = .109, p = .741, but there was an association between age and education. There was a significant difference in the age of graduates (M = 33.22, SD = 8.72) and non-graduates (M = 25.2, SD = 9.7); t(87) = -3.19, p = .002. There was no significant difference in the mean AoA per participant given by females (M = 10.6, SD = 1.74) and males (M = 10.74, SD = 2.22); t(87) = -.312, p = .756.

There was no significant difference in the mean AoA per participant given by graduates (M = 10.67, SD = 1.87) and non-graduates (M = 10.69, SD = 2.59); t(87) = .037, p = .970.

Correlations between the participant attributes of age, gender and education were checked for occurrence of collinearity. Correlations between these attributes as predictor variables and the outcome variable of mean AoA per participant were also checked.

Table 12 Summary of Pearson’s r bivariate and point biserial correlations of key demographic variables (N = 89)

		Age	Gender	Education group
Mean AoA	r	.138	.033	-.004
	p	.198	.756	.97
Age	r		.337**	.323**
	p		.001	.002

Notes: (1) Mean AoA = mean AoA per participant (2) Education group = graduate or non-graduate

There was no association between gender and education, Φ = .035, p = .741.

There were no significant correlations between mean AoA and age (r = .138, p = .198), mean AoA and gender (r_{pb} = .033, p = .756) or mean AoA and education group (r_{pb} = .004, p = .97). A significant correlation was found between age and gender (r_{pb} = .337, p = .001), older age associated with being male, and between age and education group (r_{pb} = .323, p = .002), older age associated with graduate level education.

5.5.3 Regression

Age was converted to a standardised z score before running the regression. No significant regression equation was found; ($F(3, 89) = .635$, $p = .594$), with an R^2 of .022. Estimates of effects shown in Table 13 confirm there were no significant effects of participant attributes on mean AoA ratings.

Table 13 Estimates of effects

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.851	.799		12.324	0
	Age	.035	.026	.162	1.344	.183
	Gender	-.076	.453	-.019	-.169	.866
	Education group	-.295	.601	-.056	-.490	.625

5.5.4 Item analysis

The mean AoA for each of the 1,474 agreed vowelised letter strings was calculated by summing all the ratings given by all participants for each letter string and finding the mean. This section uses correlation and multiple linear regression to analyse the data, firstly with mean *AoA* as the dependent variable and zscore(Length), zscore(Frequency) and zscore(*H*) as predictor variables. The second regression was run with the *H* statistic as the dependent variable and with zscore(Length), zscore(Frequency) and zscore(Mean AoA) as the predictor variables. *H* was used as the predictor variable for name agreement because it had provided a better fit with the data in the name agreement survey (Chapter 4).

Item characteristics were: frequency per million (M = 64.13, SD = 158.97, range .01 to 1854.84); length in letters (M = 4.17, SD = 1.03, range 2 to 8); and mean AoA (M = 10.58, SD = 2.8, range 3.76 to 19.10). Figure 13 shows the distribution of mean AoA.

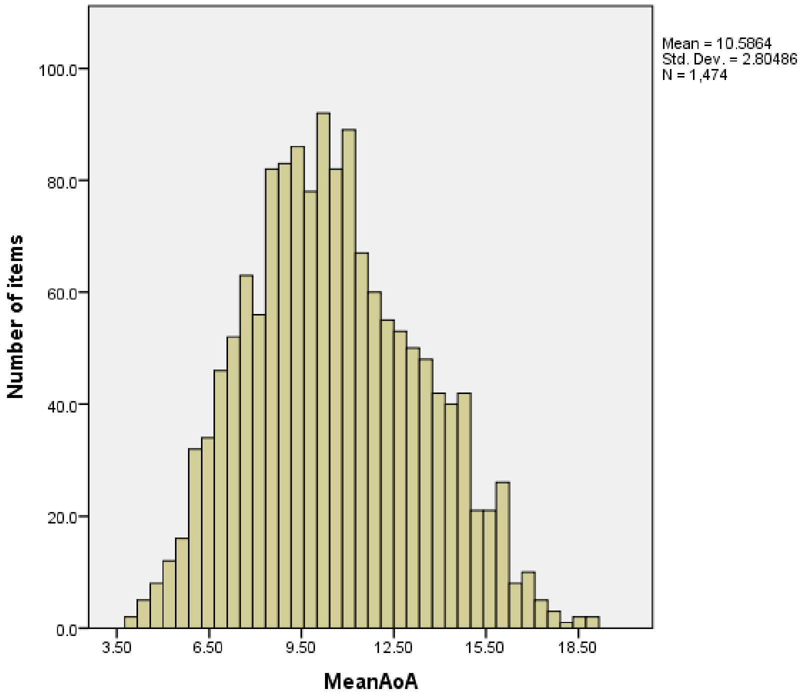


Figure 13 Distribution of mean AoA

5.5.5 Correlations

Correlations between predictors and outcome variables to be used in two regressions were also checked. AoA (mean AoA per item) and the *H*-statistic were outcome variables in the subsequent regression analyses, and correlations with their respective predictor variables were therefore checked.

Table 14 Summary of Pearson's r bivariate correlations of item variables (N = 1,474)

		H	MeanAoA	Frequency	Length	Type
Agreement	r	-.969**	.105**	-.075**	.495**	-.838**
	p	< .001	< .001	.004	< .001	< .001
H	r		-.117**	.076**	-.517**	.929**
	p		< .001	.004	< .001	< .001
MeanAoA	r			-.215**	.266**	-.137**
	p			< .001	< .001	< .001
Frequency	r				-.148**	.072**
	p				< .001	.006
Length	r					-.504**
	p					< .001

Notes: 1. Agreement = percentage name agreement 2. *H* = *H*-statistic (distribution of agreement across different correctly diacritised forms) 3. MeanAoA = mean age of acquisition per item 4. Frequency = frequency of unvowelised letter string from Aralex (Boudelaa & Marslen-Wilson, 2010) 5. Length = length in letters of unvowelised letter string 6. Type = number of different correctly diacritised forms of letter string

For mean AoA, there was a weak positive correlation with agreement (r = .105, p < .001), a higher percentage agreement on the dominant meaning of the letter string corresponding to a later age of acquisition. There was a weak negative correlation with the *H* statistic (r = -.117, p < .001), showing that more alternative forms of a letter string were associated with being learned earlier in life. There was a weak negative correlation with frequency (r = -.215, p < .001); words learned later in life were associated with lower frequency. A weak positive correlation with length (r = .266, p < .001) indicated that letter strings with more letters were associated with a later age of acquisition, and a weak negative correlation with type (r = -.137, p < .001) indicated that letter strings with more different vowelisation response were associated with an earlier age of acquisition.

Correlations between the other variables were the same as in the name agreement study described in chapter 4 and are repeated here for ease of reading. A weak negative correlation between agreement and frequency ($r = .075$, $p < .001$) indicated that an increase in frequency was associated with a decrease in percentage name agreement, while a moderate positive correlation between agreement and length ($r = .495$, $p < .001$) indicated that longer letter strings elicited greater name agreement. A strong negative correlation between agreement and type ($r = -.838$, $p < .001$) indicated that agreement decreased as the number of correctly diacritised alternatives for a homographic letter string increased.

For the *H* statistic, there was a weak positive correlation with frequency ($r = .076$, $p< .001$), a decrease in frequency corresponding to a decrease in the extent of name agreement across the number of correctly diacritised alternatives; higher frequency words had a greater number of possible alternatives. A moderate negative correlation between the *H* statistic and length ($r = -.517$, $p < .001$), indicated that shorter words had more possible alternatives. There was a strong positive correlation between *H* and type ($r = .929$, $p < .001$), with the value of *H* increasing as the number of possible alternative diacritised forms of the letter string increased.

5.5.6 Regression

Two multiple regression analyses were carried out to examine which factors influenced variation in mean AoA (by item) or the *H* statistic. The first model with mean AoA as the dependent variable had the z scores for length, frequency and *H* as the predictor variables. The second model with *H* as the dependent variables had the z scores for length, frequency and mean AoA as the predictor variables.

5.5.6.1 Regression 1: mean AoA as outcome variable

A significant regression equation was found (F(3, 1470) = 56.086, p < .001), with an R^2 of .101.

Coefficients of effects of variables are shown in Table 15.

Table 15 Survey regression model 1: Coefficients of variables

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	10.586	.069		152.819	< .001
	zscore(*H*)	.076	.081	.027	.937	.349
	zscore(Frequency)	-.504	.070	-.180	-7.188	< .001
	zscore(Length)	.710	.082	.253	8.700	< .001

The estimated coefficient for length implies that AoA changes by .710 for unit increase in standardized length: longer unvowelised letter strings tend to have been learned later. Zscore(Length) significantly predicted AoA, β = .710, t = 8.7, p < .001. The estimated coefficient for frequency implies that AoA changes by -.504 for unit increase in standardised frequency. For every unit decrease in frequency, a unit increase in mean AoA is expected. Zscore(Frequency) was significant, β = -.504, t = .937, p = .349.

5.5.6.2 Regression 2: *H* as outcome variable

A second regression was run using the *H* statistic as the dependent variable with Z scores for length, frequency and mean AoA as the predictor variables, in order to determine whether mean AoA or the *H* statistics would give a better fit between model and data. A significant region equation was found (F(3,1470) = 178.795, p < .001) with

R^2 = .266. Table 16 shows the coefficients of variables. Zscore(Length) was the only significant predictor variable of *H*, β = -.4, t = -22.435, p < .001. Zscore(Frequency) was not significant, β = .003, t = .143, p = .886. Zscore (MeanAoA) was not significant, β = .017, t = .937, p = .349.

Table 16 Coefficients of variables: model 2

Model		B	Std. Error	Beta	t	Sig.
2	(Constant)	.771	.017		45.078	<.001
	Zscore(MeanAoA)	.017	.018	.022	.937	.349
	Zscore(Frequency)	.003	.018	.003	.143	.886
	Zscore(Length)	-.400	.018	-.522	-22.435	<.001

a. Dependent Variable: H

Comparison of the R^2 values for the two models indicates that the second model, with AoA as one of the predictor variables and *H* as the outcome variable provides a better fit with the data (R^2 = .266 for model with *H* as outcome variable, compared with R^2 = .101 for model with AoA as outcome variable). Length of letter strings in letters as a predictor variable is common to both models.

5.6 Discussion

A significant effect of length was found in both regression analyses; length of unvowelised strings in letters predicted both AoA and *H*. However, frequency significantly predicted only AoA (the first model). The correlation between AoA and frequency was similarly found with standardised reaction times in the English lexicon project (Kuperman et al., 2011). The level of correlation between AoA and frequency in the present study was lower than that found in a number of studies reviewed by

Brysbaert and Ghyselinck (2006). A possible explanation is that the present study used AoA ratings for the vowelised form of the letter string with highest name agreement for the 60% of items which were homographic, but used the frequency for the unvowelised letter string. This frequency would be larger for the target items with more than one alternative vowelisation and would not reflect the frequency of the agreed vowelised form. Shorter words were learned earlier in life and had more possible vowelisations and pronunciations (longer words learned later had fewer). Multiple vowelisations were mapped onto a single orthographic form, namely the unvowelised letter string, during the years of education.

One difference between the present AoA study and Kuperman et al. (2011) was that less influence of age, gender or education was found on how AoA was rated. This finding may have been influenced by the lack of detail regarding participants' education level which, in the present study, was simply divided into graduate or non-graduate In contrast to the study by Kuperman et al. (2011) who, in English, were able to compare their AoA estimates with those of other studies, the present study has no such benchmark against which to compare its findings. Moreover, Arabic speakers will usually have learned MSA words for everyday items such as 'bread' and 'water' later than their colloquial equivalents and the present study avoided vocabulary items which were most likely to have been acquired before primary school. This gives a reason for what may appear older AoA ratings. Distribution of AoA ratings shown in Figures 8 to 10 illustrate that distributions were broadly normal

Based on the review of studies by Juhasz (2005), the effect of AoA should be seen in the word naming and lexical decision experiments in the present study. This is in accordance with the findings in English (Balota et al., 2001) that AoA predicted response times in both tasks. To date, research in Semitic languages has concentrated more on children’s actual AoA (objective AoA) in the context of the development of literacy. Hence, in MSA there is no list of vowelised forms of unvowelised letter strings which includes nouns and verbs, and which can be used in further experiments. When Cortese and Khanna (2008) published AoA ratings for 3,000 monosyllabic English words, the variable then became available for use in later analyses of these words. Thus, the findings of the present survey may be the starting point for investigating AoA norms in skilled adult Arabic readers. The effect of length is expected to be seen in the word naming and lexical decision experiments.

Due to high correlations between AoA and other variables, the effect of AoA has been questioned, although it is argued that, empirically, learning a word (AoA) is a behavioural event whereas using the word over time (frequency) is not (Zevin & Seidenberg, 2002). Brysbaert & Ghyselinck (2006) have proposed two types of AoA effect, one dependent on frequency and the other not; the frequency-independent effect is hypothesized to result from competition at the level of meaning. Support for the role of semantics in explaining has been provided by Cortese & Khanna (2007) who found a larger effect of AoA in lexical decision than in word naming tasks.

The AoA effect is relatively difficult to explain in models of word recognition and naming. Dual Route Cascaded theory (Coltheart et al., 2001) does not take the process of language acquisition into account. Connectionist models have not fully replicated the process of learning to read, despite the work of, among others, Ellis and Lambon Ralph (2000) and Monaghan and Ellis (2010). Ellis and Lambon Ralph (2000) designed training of their network model to include earlier- and later-acquired items. They found that items acquired early had an advantage over those acquired later and proposed this was due to the network being more flexible in its earlier stages. This network could therefore be shaped to a greater extent by words acquired early, when it was still being formed. The developmental model produced by Monaghan and Ellis (2010) demonstrated that the effect of AoA was distinct from the effect of frequency. Words learned later by the network had a smaller impact on the connections in the network, and weights on the connections were adjusted less. Frost (2012) has proposed that any model of reading must take into account the differences between languages in terms of language development. The type of model involving the developmental aspect of reading produced by Monaghan and Ellis (2010) could be a useful starting point for doing the same in Arabic, either or both the colloquial and Modern Standard versions of the language. This may be an important consideration in Arabic where shorter and higher frequency words are associated with a higher degree of homography and therefore unvowelised homographic letter strings may benefit from an AoA advantage. A developmental model would enable this to be tested. More immediately, the present AoA ratings are the first to be available for adult native Arabic speakers and can be used to assist development of stimuli, by adding further variables, and stimulus selection in subsequent research.

Chapter 6 WORD NAMING

6.1 Introduction

This chapter examines the effect of name agreement, Age of Acquisition (AoA), length and frequency effects in how they influence responses to unvowelized letter strings in Arabic, including homographic and non-homographic words. The chapter begins with hypotheses for the present experiment based on the literature review, followed by selected information about previous research in English, Hebrew and Arabic regarding ambiguity and homographs. This is followed by information about the participants, materials and procedures used before correlation and mixed effects regression analyses of the data are presented. The chapter concludes with a discussion of the findings and the implications for how unvowelised letter strings are processed in Arabic and for the organisation of the mental lexicon.

6.2 Objective and hypotheses

This experiment was carried out to investigate how native speakers name isolated Arabic words when they are presented as unvowelised letter strings, some of which are homographic and others are non-homographic. It focuses on reading a list of unvowelised letter strings which contains a high proportion of heterophonic homographs, letter strings which can be vowelised and pronounced in more than one way (887 of 1,474 items or 60.17%).

In picture naming, name agreement has been shown to predict response times in a variety of languages including British English (Ellis & Morrison, 1998), Belgian Dutch (Severens et al., 2005), French (Bonin et al., 2002), Italian (Dell'Acqua et al., 2000) Spanish (Cuetos et al., 1999) and Welsh (Barry et al., 1997). Pictures with one dominant response are named more rapidly than those to which there are many alternative responses (Alario et al., 2004; Barry et al., 1997). In picture naming, the name agreement effect has been found to be a stronger predictor of response times than the word frequency effect (Alario et al., 2004).

The first hypothesis is therefore that (1) response times will be shorter for non-items, i.e.non-homographic items, which have only one possible pronunciation.

It is well known that, in word naming, frequently encountered words elicit faster responses than those encountered less often (e.g., Balota & Chumbley, 1984; Balota & Spieler, 1999). In Hebrew, Bentin and Frost (1987) found that high frequency unvowelised letter strings were responded to more quickly than low frequency unvowelised letter strings, and that the word frequency effect was significant for ambiguous as well as unambiguous words. The second hypothesis is therefore that (2) high frequency unvowelised letter strings will be named faster than low frequency unvowelised letter strings, whether they are homographic or not

Word naming performance in alphabetic languages is affected by numerous variables including length (Balota et al., 2004). The length of a word is said to affect the time taken to pronounce it (Ferrand et al., 2011). Although some studies have found no effect of length, six of the nine word naming studies in European languages mentioned by New et al. (2006) found that longer words elicited slower responses. There is no directly relevant literature regarding Semitic languages or abjads. The third hypothesis is therefore that (3) longer unvowelised letter strings will have longer response times, whether they are homographic or non-homographic.

It is argued that there are two types of AoA, one related to frequency and the other a semantic effect (Brysbaert & Ghyselinck, 2006). The effect of AoA can be distinguished from that of frequency (Cortese & Khanna, 2007, 2008; Juhasz, 2005; Velan & Frost, 2011). AoA predicts naming latencies (Balota et al., 2004; Juhasz,

2005) and early acquired words are responded to more quickly (Morrison & Ellis, 2000; Ferrand et al., 2011). The fourth hypothesis is therefore that (4) early-acquired unvowelised letter strings will be named faster than those acquired later, whether they are homographic or non-homographic.

6.3 Background

Ambiguity can take different forms and the prevalent type of ambiguity can vary between languages. In English, ambiguity is typically related to meaning; more than 80% of frequently encountered words have multiple meanings (Rodd et al., 2002). Rodd et al. (2002) differentiated between words sharing an orthography which happen to have unrelated meanings, such as ‘chap’ (a man, or a sore place on the skin), and those which have many related senses, for example ‘roll’, a piece of bread or a movement or the metaphorical ‘on a roll’, all of which are associated with movement, even though rolling dough for the bread roll or the fact it can be rolled over (unlike a loaf) is further removed. In a series of three lexical decision experiments, they compared response times to ambiguous words with few and many senses, and unambiguous words with few and many senses. Words with many related senses were found to elicit faster responses than those with few but different meanings. They hypothesized that semantic relationships among many senses could assist word recognition, whereas multiple different meanings could delay response times because of competition among the meanings. In Hebrew, Peleg et al. (2007) indicated that the dominant meaning of isolated homographic letter strings would be the only meaning accessed if presented without a context that provides clues; ambiguity resides in the orthography as well as in the meaning.

Rodd (2004) tested whether related and unrelated meanings of word elicited differing response times, using a word naming experiment with a 2 x 2 factorial design of ambiguity (many versus few senses) and consistency of spelling (consistent versus. inconsistent), followed by a second word naming experiment involving only inconsistent items and a different group of participants. Comparison of latencies for the inconsistent items in both experiments showed that in the second experiment where only inconsistent items were investigated, latencies were longer than in the first experiment, and a significant advantage was found for words with many senses. Rodd proposed that removal of consistent items which were easier to access may have compelled participants to use a slower route to name the inconsistent items. Rodd concluded that in reading aloud, skilled readers did access meaning.

In printed Hebrew text, ambiguity originates from an additional source, namely the omission of short vowels. Short vowels are represented in Hebrew in a similar way to Arabic, although the unvowelised text is referred to as 'unpointed' in Hebrew rather than as 'without diacritics' or 'unvowelised' as used to describe the same phenomenon in Arabic. The omission of short vowels leads to a higher number of possible ways of pronouncing an individual letter string as well as different meanings and senses (Bentin & Frost, 1987). It is possible for a letter string to be ambiguous in pronunciation, but also possible for a single pronunciation to have different and unrelated meanings. Comparisons of word naming latencies in Hebrew, English and Serbo-Croatian have indicated that naming of unvowelised Hebrew letter strings involved selection of one of several possible pronunciations (Frost et al., 1987). Frost and Bentin (1987) aimed to test the orthographical depth hypothesis using word naming to investigate whether lexical search took longer as orthographic depth increased, and using lexical decision response times as a benchmark for comparison of corresponding results of word naming. They found that high frequency letter strings were identified faster, even when they had more possible pronunciations.

Further experiments used lists of ambiguous and unambiguous items, with and without vowels, together with nonwords, in order to compare the effect of frequency on naming the different types of word (Bentin & Frost, 1987). Frequencies were rated as high or low on a scale of one to five by 50 undergraduates. Results of paired comparisons of high and low frequency alternatives, of both vowelised and unvowelised letter strings, indicated that high frequency words were named faster than low frequency words and that unvowelised letter strings were named as quickly as vowelised words. Based on mean response times (RTs) calculated for word and subject, subjects chose the high frequency alternative of the two unvowelised letter strings more often. Bentin and Frost (1987) proposed that word naming in Hebrew required disambiguation of both meaning and phonology and that the ambiguity was mainly resolved by the relative frequencies of the competing alternatives. They proposed that initial lexical access was followed by a race between the pronunciations and meanings of the low- and high-frequency alternatives. However, their analysis did not seek to include the effects of other variables which could have affected latencies, such as length and neighbourhood density (Spieler & Balota, 2000). Thus, other factors could have been involved in the findings, which could in turn have led to other explanations. In particular, neighbourhood density has been shown to affect response times (e.g. Andrews, 1997).

Printed Arabic text, like Hebrew, contains many homographic words due to the omission of the diacritics representing short vowels. However, there are few studies of the effects of ambiguity on the process of word naming in Arabic. Seraye (2004) found that frequency was the only significant predictor of response times, possibly because too many other variables were included in a list of 124 stimuli. 90 items were allocated to three groups of 30 items under different reading conditions: with short vowels, with short vowels and a wrongly placed long vowel marker (the shaddah), and variation in frequency for both homographic and non-homographic letter strings. The stimulus set included several parts of speech and affixes in order to vary the classes of word. Word length was also varied. Critically, any of the possible alternatives for a homograph was considered a legal response, so the study revealed little about the effect of homographs or of the variables that could affect their processing. Longer latencies for vowelised low-frequency words were explained in terms of the Dual Route Cascaded theory (Coltheart et al., 2001), stating that participants had to use the phonological route to access the mental lexicon.

Other studies in Arabic involving homographs have focused primarily on the means of disambiguation: the role of vowels (e.g. Abu-Rabia, 2001), the role of diacritics (e.g. Hermena, Drieghe, Hellmuth, & Liversedge, 2015), role of sentence context (e.g. Abu-Rabia, 1997a), the impact of diglossia (e.g. Saiegh-Haddad, 2004) and the processing of morphological units (e.g. Boudelaa, 2013). These researchers have offered different explanations of why vowels, context, diglossia and Arabic morphology help or hinder disambiguation. Vowels make the pronunciation and meaning clear, whereas diacritics are considered to lead to perceptual overload which slows the processes of disambiguation. Context provides clues for determining the appropriate pronunciation and meaning of an unvowelised letter string, while identification of root and word pattern morphemes can give access to pronunciation and meaning although for isolated words it will not by itself determine which of the several possibilities is appropriate. The present study focuses more directly on heterophonic homographs. In response to the assertion that frequency is the major predictor of latencies in Arabic, which it is in English (e.g. Balota et al., 2006), as indicated in the studies described here, the influence of name agreement and AoA is examined.

When a word or letter string is ambiguous, different individuals may understand it differently. One measure of the extent to which words are understood differently by different individuals is name agreement. Name agreement has been shown to have demonstrable effects in picture naming tasks; pictures with one dominant response are named faster and more accurately than pictures to which there are many different multiple responses (Barry et al., 1997). Effects of name agreement as a predictor of response times in picture naming have been found in various languages including British English (Ellis & Morrison, 1998), Spanish (Cuetos et al., 1999) and Welsh (Barry et al., 1997). These three studies all reported that name agreement strongly predicted response times in picture naming. It is suggested that higher name agreement results in activation of fewer lexical candidates, in contrast to lower name agreement where there are more candidates and therefore more time is needed to make a selection (Alario et al., 2004). Unvowelised letter strings with lower name agreement would indicate a greater number of lexical candidates and, based on the findings of Alario et al. (2004) would be expected to be responded to more slowly.

Interest in the influence of AoA in different languages is accompanied by findings that its strength as a predictor of latencies varies by language. In English, Cortese and Khanna (2008) found that AoA as a predictor variable was additional to those already identified by Balota et al. (2004), as explained in section 3.2.5. Earlier AoA is associated with shorter latencies in English (Brysbaert & Cortese, 2010; in Italian (Wilson, Ellis, & Burani, 2012), Japanese (Havelka & Tomita, 2006) and Spanish (Davies, Barbón & Cuetos, 2013). Moreover, picture naming experiments have confirmed that AoA influence can be detected in perceptual, semantic and phonological processing, prelexical and postlexical operations (Navarrete, Pastore, Valentini, & Peressotti, 2013).

In Hebrew and Arabic, support has been found for the Orthographic Depth Hypothesis (Katz & Frost, 1992) which proposes that word recognition is achieved more easily in shallower orthographies. The perceptual load in Semitic languages leads to slower word recognition and naming because of the cursive script and diacritics (e.g. Ibrahim et al., 2002). Ibrahim (2013) carried out a study with 75 children in Grade 8, using 50 words of each of unvowelised, vowelised and pseudo-words in lists which increased in difficulty. Children were asked to read them as accurately and quickly as they could. Ibrahim (2013) found that unvowelised letter strings were read aloud faster and with greater accuracy than the vowelised forms and proposed this was due their reliance on visual clues since they had not read vowelised text at school for several years.

6.4 Method

6.4.1 Participants

A total of 38 adult native speakers of Arabic (19 males and 19 females, mean age 28.08 years, standard deviation 7.265, age range: 18-43 years) were recruited, 26 from the population of Saudi international students and their spouses in the UK and a further 12 in Riyadh, Saudi Arabia. 20 participants were educated to bachelor's degree level or above, while the remaining 18 had completed their education before obtaining a bachelor's degree. Their mean number of hours spent reading Modern Standard Arabic (MSA) per day was 1.7 hours (SD of 1.07, range: 0.5 to 5 hours). Participants received 15 pounds for travel expenses. All were native speakers and had normal or corrected-to-normal vision. No participant reported having been diagnosed with a learning difficulty or language impairment.

Ethical approval was obtained from Lancaster University as well as Imam Muhammad University in Riyadh, Saudi Arabia (see Appendix C).

6.4.2 Materials

Participants were asked to read aloud the 1,474 unvowelised letter strings previously presented in the name agreement and AoA studies. Frequency estimates for the unvowelised letter strings as they appeared in printed Arabic text, expressed as frequency per million, were taken from the Aralex (Boudelaa & Marslen-Wilson, 2010) corpus of 40 million MSA words, drawn from a variety of online newspapers.

As reported previously, the normative dominant pronunciation, and corresponding meaning, of the 1474 letter strings was determined in a name agreement study (see Chapter 4). In that study, it was found that 587 strings elicited 100% name agreement, evoking one spelling and one pronunciation only, according to the responses produced by participants in the agreement study. Each of the other 887 strings evoked more than one legitimate spelling (vowelisation) and pronunciation. The modal (most frequent, dominant) pronunciation, of those recorded for each string, was taken to be the correct pronunciation for that string. In total, the 1,474 letter strings were found to correspond to 1,064 nouns and 410 verbs, counting the meaning of the modal response, where strings evoked multiple different vowelisations.

Word lengths in the present study were defined as the number of letters in the unvowelised form found in printed text and were counted individually by the researcher. The target words corresponding to the strings were 2 to 8 letters in length (without diacritics, as seen on screen). As described in Chapter 5, the unvowelised letter strings words were divided into three groups, two of 500 items and one of 474 items, and were given AoA ratings in actual years by three groups of different participants (89 in total. AoA ratings were collected only for the agreed vowelised forms of the words, the modal responses. Table 17 summarises item characteristics, providing the mean, SD, minimum and maximum values, and range for each variable.

Table 17 Summary of item characteristics (N = 1,474)

Variable	Mean	SD	Min.	Max.	Range
agreement	.780	.231	.210	1.000	.79
H	.721	.752	.001	2.940	2.94
Type	2.400	1.733	1	11	10
meanAoA	10.486	2.805	3.800	19.100	15.3
Frequency	64.1	159.0	.010	2854.84	2854.83
Length	4.170	1.037	2	8	6

Figures 14 to 16 show the distributions of frequency, length and AoA.

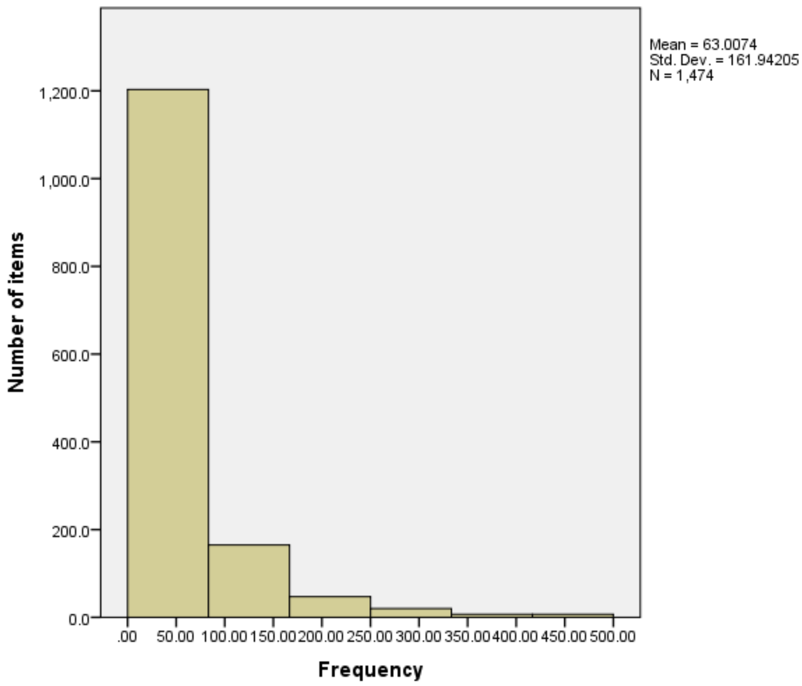


Figure 14 Distribution of letter string frequency

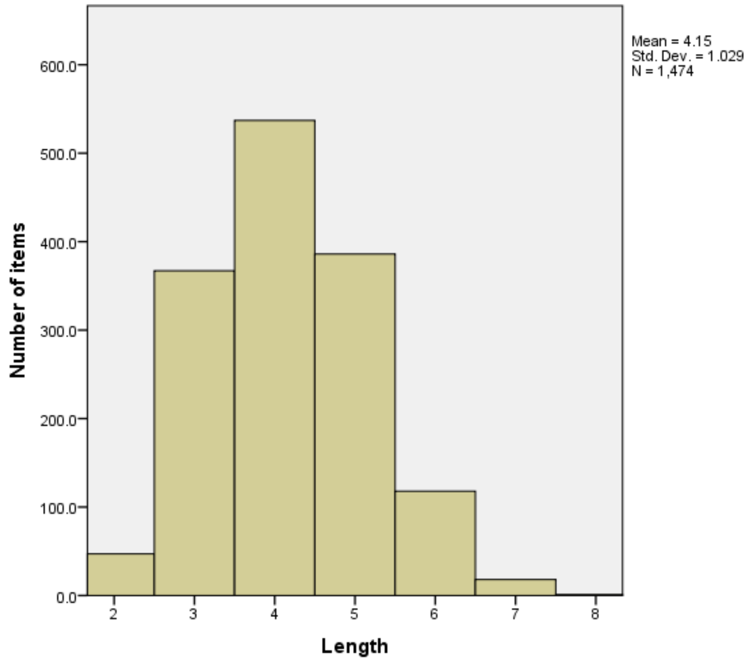


Figure 15 Distribution of letter string lengths

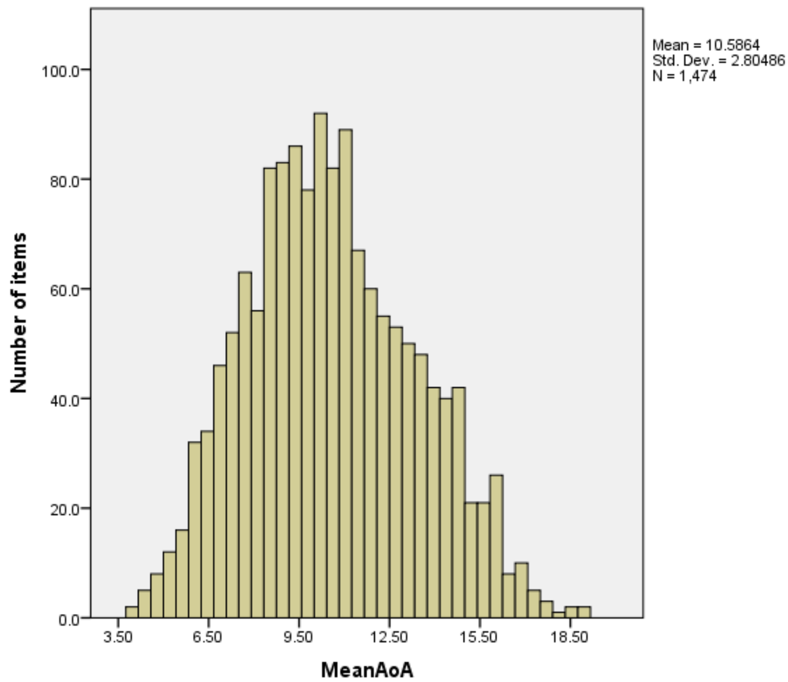


Figure 16 Distribution of mean AoA of letter strings

Neighbourhood density values were not entered into the analysis of naming responses. This is because the calculation of neighbourhood density for unvowelised letter strings in Arabic is problematic. The only available information on neighbourhood density corresponds to the neighbourhoods of fully vowelised Classical Arabic rather than MSA, with estimates produced for the Qur'anic Lexicon Project (Binte Faisal et al., 2015). These cannot apply to unvowelised letter strings which will have different densities if they have unrelated meanings, as some of them do. Neighbourhood densities for unvowelised homographs would not be comparable with measures in non-Semitic languages because changing one letter in the root would generate a word which could be very distant in terms of semantics and morphology because the root carries the core meaning for a whole family of words. A different root carries a different core meaning. This would not produce an accurate measure of orthographic neighbourhood density as it is understood in English because it would bring in morphological density pertaining to a different family of words. Similar problems pertained to phonological and morphological neighbourhood density (for a full explanation see section 3.2.3). For this reason, in order to provide some measure of density, the number of types (different vowelised words) for each heterophonic homographic word, as identified by participants in the name agreement study, was used in place of neighbourhood density.

To advertise the opportunity to participate in the study, information sheets were distributed among Saudi students and their dependants in the UK. The information sheet described the purpose of the experiments, the incentive offered, the right to withdraw and the methods to be used if the participant wished to take part. When the participant agreed to take part, they were first asked to sign a consent form, then asked to give some basic personal information (age, gender, education level, the number of hours spent reading MSA per day), before completing the reading task.

6.4.3 Apparatus and procedure

Subjects were instructed to read aloud each of the 1,474 unvowelised letter strings in the stimulus set. Stimuli were presented one at a time on screen. The 1,474 trials were divided into ten blocks, with strings assigned to seven subsets of 148 items and three of 146 items. Participants were allowed to take a break between each block of trials, resuming when they were ready. On average, participants completed the task in approximately two hours. The presentation of items within blocks, and of blocks within a session, were randomized for each participant.

Participants were each tested individually in a quiet room. They were instructed to read each word in standard Arabic, as quickly and accurately as possible, and to avoid making a sound before reading the word aloud. They were instructed to read each word in the form that came to mind as all words were presented unvowelised and without diacritics. Instructions appeared on the computer screen where they were seated prior to the start of the task and were read aloud by the experimenter.

Participants had to complete the task in a single session. Each participant was given 20 practice trials, consisting of letter strings of similar length to those used in the actual trials, after which he or she was given the opportunity to ask any questions. For each trial, each participant was seated 50 cm from the computer screen. A fixation point was presented at the centre of the screen for 500 ms, followed by presentation of the unvowelised letter string in black size 16 font at the centre of a light grey screen, where it remained for 2,000 ms. A blank screen then appeared for 500ms.

The DMDX application (Forster & Forster, 2003) was used for presentation of the stimuli and to record the spoken response to disk through a headset microphone. Response latencies were extracted from the responses recorded on the sound file using the CheckVocal program (Protopapas, 2007). CheckVocal was used to ensure that no sounds unrelated to the naming of the word were included in the latency.

6.5 Results

6.5.1 Data preparation

A total of 56,012 responses were recorded. Only response times (RTs) to correct responses were analysed. Responses were classified as correct if they consisted of the modal response for the target letter string, identified in the name agreement survey. Data from 6,080 trials were not analysed because the responses were excluded as incorrect, according to this classification scheme, or because they were not completed before the end of the 2,000ms response interval, or because no response was made. Exclusions represented 10.85% of the total number of responses, with 49,932 data items remaining for analysis.

6.5.2 Data analysis

Correct response latencies were analysed using correlation and linear mixed-effects models, in IBM SPSS (version 22). Uncontrollable variability by item and subject, for example the mood of a participant at the time of the trial, were incorporated by the inclusion of random effects. Mixed effects models have been shown to detect non-zero effects more reliably than other kinds of regression (Baayen, Davidson, & Bates, 2008). Raw response times were therefore used in the mixed effects modelling. For the purpose of obtaining correlations between variables, mean response times per item were calculated. The mean response time for all items was 686.67 ms (SD = 63.23). Response times ranged from 542.9 to 984.8 ms. The distribution of response times is shown in Figure 17.

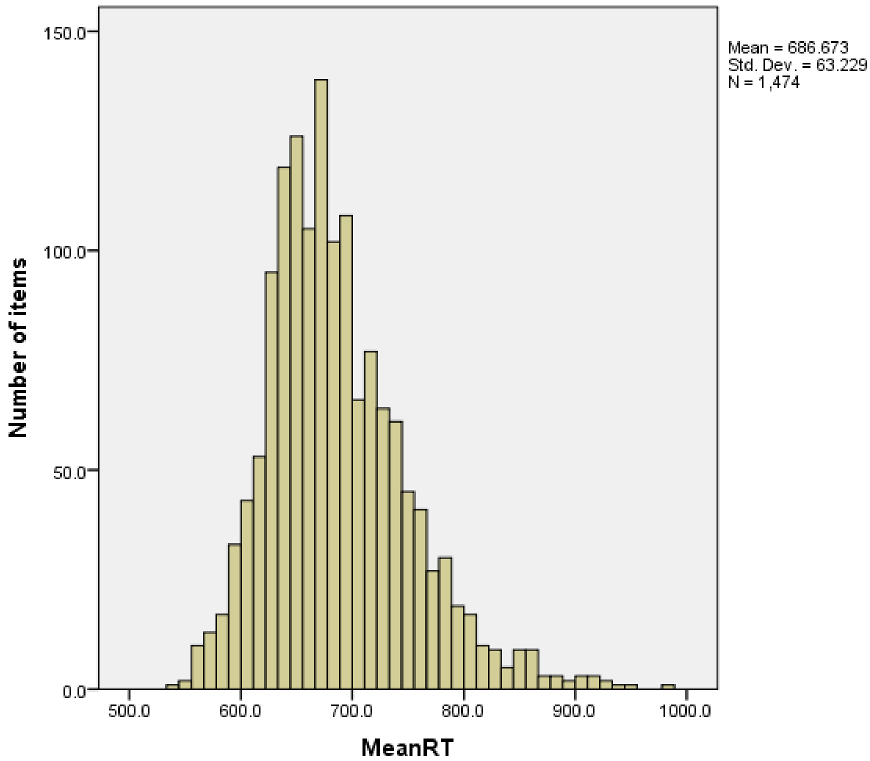


Figure 17 Distribution of mean response times for all items: word naming

6.5.3 Correlations

Correlations were calculated to identify which predictor variables had a significant relationship with each other and to highlight possible occurrences of collinearity. They were also used to check the relationships among alternative measures of the extent of agreement about the number of possible vowelisations of letter strings: name agreement, type and the *H* statistic (which takes into account the number of different uses of correct vowelisations of the letter string). Critically, they were used to examine the relationships between mean response times (outcome variable) and the predictor variables.

Table 18 Summary of Pearson's r bivariate correlations between key psycholinguistic variables (N = 1,474 in all comparisons)

		H	Agree-ment	Type	AoA	Frequency	Length
MeanRT	r	.202**	-.194**	.178**	.355**	-.105**	-.053*
	p	< .001	< .001	< .001	< .001	< .001	.042
H	r		-.968**	.929**	-.118**	.076**	-.516**
	p		< .001	< .001	< .001	.004	< .001
Agreement	r			-.838**	.104**	-.075**	.495**
	p			< .001	< .001	.004	< .001
Type	r				-.137**	.072**	-.504**
	p				< .001	.006	< .001
AoA	r					-.215**	.265**
	p					< .001	< .001
Frequency	r						-.148**
	p						< .001

Notes: (1) Mean RT = mean response time. (2) *H* = *H* statistic (measure of distribution of name agreement across all alternative diacritised forms) (3) Agreement = percentage name agreement among participants (4) Type = number of different alternative forms of homograph (1 to 11) (5) AoA = participant-rated mean age of acquisition by item (6) frequency from Aralex (Boudelaa & Marslen, 2010) (7) Length = length of letter string in letters

There was a weak positive correlation between mean RT and the *H* statistic, longer response times corresponding to a greater number of possible pronunciations (r = .202, p < .001), a very weak positive correlation between mean RT and type, longer response times associated with unvowelised letter strings with more possible vowelised forms (r = .178, p < .001) and a weak positive correlation between mean RT and AoA, indicating than later-acquired words elicited slower response times (r = .355, p < .001).

There was a weak negative correlation between mean RT and agreement, where higher percentages of agreement were associated with shorter latencies ($r = -.194$, $p < .001$), a weak negative correlation between mean RT and frequency, showing that higher frequencies were associated with shorter response times ($r = -.105$, $p < .001$) and a weak negative correlation between mean RT and length, longer words being associated with shorter latencies ($r = -.053$, $p = .042$).

There was a strong negative correlation between agreement and H, indicating that percentages of name agreement rose as values of H decreased ($r = -.969$, $p < .001$), a weak positive correlation between agreement and AoA, indicating that agreement increased with AoA ($r = .105$, $p < .001$), a weak negative correlation between agreement and frequency, pointing to higher frequency words having lower percentage name agreement ($r = -.075$, $p < .001$), a moderate positive correlation between agreement and length, where there was greater percentage agreement for longer words ($r = .495$, $p < .001$), and a strong negative correlation between agreement and type, indicating that percentage agreement was lower for letter strings which were associated with more alternative vowelisations ($r = -.838$, $p < .001$).

For the *H* statistic, there was a weak negative correlation with AoA, lower values of *H* being associated with later-acquired words ($r = -.117$, $p < .001$), a weak positive correlation with frequency, indicating that higher values of *H* (higher name agreement) were associated with higher frequency ($r = .076$, $p < .001$), a moderate negative correlation with length ($r = -.517$, $p < .001$), and a strong positive correlation with type ($r = .929$, $p < .001$).

For mean AoA, there was a weak negative correlation with frequency, earlier AOA being associated with higher frequency ($r = -.215$, $p < .001$), a weak positive correlation with length, indicating longer words were associated with later AoA ($r =$

.266, $p < .001$) and a weak negative correlation with type, with earlier-acquired items having elicited more alternative vowelisations in the name agreement survey ($r = -.137$, $p < .001$). There was a weak negative correlation between frequency and length, such that higher frequency letter strings were associated with shorter lengths of letter strings ($r = -.148$, $p < .001$), and a very weak positive correlation between frequency and type, higher frequency being associated with a greater number of alternative forms of a letter string ($r = .072$, $p < .001$). Finally, there was a moderate negative correlation between frequency and length, where longer words were associated with lower frequency ($r = -.504$, $p < .001$).

6.5.4 Mixed effects regression

Two mixed effects models were fitted to examine the effects of critical psycholinguistic variables on Arabic word naming latencies. In both models, random effects of participants and of items on intercepts were specified, to take into account random differences between participants or between target letter strings in the average latency of the responses recorded. In both models, fixed effects were specified to estimate the effects of: the orthographic frequency of the letter string; the length of the unvowelised string in letters; the AoA of the modal (dominant meaning, pronunciation) of the string; and the diversity of vowelisation responses to the string. In one model, the name agreement statistic was used to capture the effect of vowelisation diversity. Name agreement calculates the percentage of participants agreeing with the modal form. The results of this model are reported first. In the second model, vowelisation diversity was captured using the *H* statistic, which takes into account the number of different uses of correct vowelisations of the letter string. These measures were examined separately to establish which would be more useful as a predictor variable to provide a better fit with the data. The variable ‘type’ was not used because it was included in the calculation of the *H* statistic.

The models included terms specifying the effects of participant attributes, including age, gender and education in order to examine whether participant characteristics affected the response times, and to be able to distinguish such effects from the effects of the psycholinguistics variables.

The models also included terms specified to estimate the effects of interactions between the effects of the critical psycholinguistic variables. It was expected that there would, potentially, be interactions between the effects of the critical psycholinguistic variables. It was anticipated that the effect of agreement would potentially interact with the effects of length (agreement x length interaction), frequency (agreement x frequency interaction) or AoA (agreement x AoA interaction). It was also predicted that the effect of length would interact with the effects of frequency (length x frequency interaction), or AoA (length x AoA interaction). It was predicted that the effects of frequency and AoA would interact (frequency x AoA interaction).

All continuous numeric variables were standardized to z scores before inclusion in models. Interaction terms were computed as the multiplicative product of lower-order terms, e.g., the term corresponding to the interaction between the frequency and AoA effects was calculated as the product of the standardized AoA and frequency variables.

All possible variables were entered as dimensions of the first model predicting response times.

6.5.4.1 Mixed effects regression model 1

The first model was run with agreement, frequency, length, AoA and the interactions between these different variables as predictor variables. Participant attributes were also included to ascertain whether they introduced bias into the result.

All possible variables were entered as dimensions of the first model predicting response times. The estimates of fixed effects are presented in Table 19. Main effects are reported, together with interaction effects. Non-significant effects are also noted.

Table 19 Word naming model 1: Estimates of fixed effects

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Fixed effects							
Intercept	687.358	23.267	38.355	29.542	< .001	640.270	734.445
[GenderCode=0]	-50.411	26.863	38.000	-1.877	.068	-104.792	3.970
[GenderCode=1]	.000		.	.	.	.	.
[QualificationCode=0]	50.982	28.786	38.000	1.771	.085	-7.291	109.255
[QualificationCode=1]	.000		.	.	.	.	.
zAge	15.445	15.001	38.000	1.030	.310	-14.922	45.813
zAgreement	-14.694	1.710	1427.673	-8.592	< .001	-18.049	-11.339
zAoA	22.885	1.578	1401.560	14.506	< .001	19.790	25.980
zFrequency	-11.859	2.423	1378.984	-4.894	< .001	-16.613	-7.105
zLength	-3.272	1.759	1385.605	-1.860	.063	-6.723	.179
zAGxzL	-6.224	1.737	1423.371	-3.583	< .001	-9.632	-2.817
zAGxzF	-2.164	1.553	1409.773	-1.393	.164	-5.211	.883
zAGxzAoA	.933	1.764	1415.325	.529	.597	-2.527	4.393
zLxzF	-4.902	2.343	1372.355	-2.092	.037	-9.499	-.305
zLxzAoA	-.224	1.629	1382.278	-.138	.891	-3.420	2.972
zFxzAoA	-5.151	2.220	1399.941	-2.320	.020	-9.506	-.795
Random effects							
				Variance	Std. Error		
Item (intercept)				2394.63	126.459		
Subject (intercept)				6226.88	112.993		
Residual				19662.82	1432.01		

Number of observations: 49,932; groups = 1,474 items and 38 subjects

Notes: (1) GenderCode = 1 male 0 female (2) QualificationCode = 1 graduate 0 non-graduate (3) ZAge = age, standardised score (4) ZAgreement = percentage agreement among participants, standardised score (5) ZAoA = participant-rated mean age of acquisition by item, standardised score (6) ZFrequency = frequency of unvowelised form taken from Aralex (Boudelaa & Marslen-Wilson, 2010), standardised score (7) ZLength = length of unvowelised letter string in letters, standardised score (8) ZAG = abbreviation of percentage agreement among participants, standardised score (9) ZL = abbreviation of length, standardised score (10) ZF = abbreviation of frequency, standardised score

It can be seen in Table 19 that participant attributes did not significantly affect response times: age (SE = 15.001, t = 1.0, p = .310), education (SE = 28.786, t = 1.7, p = .085), gender (SE = 26.863, t = -1.8, p = .068). As can be seen in Table 19, three of the main effects reached significance in accounting for the variance in the reaction-time data; agreement (SE = 1.710, t = -8.6, p < .001), AoA (SE = 1.578, t = 14.5, p < .001) and frequency (SE = 2.423, t = -4.9, p < .001). However, all of these were also influenced by interactions with other variables and so main effects and effects of interactions are discussed together.

If all other effects are zero, then for unit increase in name agreement, RT decreases by 14.7ms. A significant interaction between agreement and length (SE = 1.737, t = -3.6, p < .001) shows that the effect of name agreement is different for letter strings of different lengths. As the lengths of letter strings increase the size of the name agreement effect increases by -6.2. Interactions between name agreement and frequency (SE = 1.553, t = -1.4, p = .164) and between name agreement and AoA (SE = 1.764, t = .5, p = .597) were not significant.

If all other effects are zero, then for unit increase in AoA, RT increases by 22.9ms. An interaction between AoA and frequency (SE = 2.220, t = -2.3, p = .020) indicates that the effect of AoA is different for letter strings of different frequencies. As the frequencies of letter strings increase, the size of the AoA effect increases by -5.1. Interactions between AoA and name agreement and between AoA and length (SE = 1.629, t = -.1, p = .891) were not significant.

If all other effects are zero, then for unit increase in frequency, RT decreases by 11.9ms. An interaction between frequency and length (SE = 2.343, t = -2.1, p = .037) shows that the effect of frequency differs for letter strings of different lengths. As the lengths of letter strings increase the size of the frequency effect increases by -4.9. An interaction between frequency and AoA (SE = 2.220, t = -2.3, p = .020) shows that the effect of frequency differs for words acquired at different ages. As the AoA increases, the size of the frequency effect increases by -5.1. Interaction between frequency and name agreement and frequency and AoA, as previously stated, was not significant.

The effect of length alone is not significant (SE = 1.759, t =-1.860, p = .063). However, an interaction between length and name agreement (SE = 1.737, t = -3.6, p < .001) shows that the effect of name agreement differs for letter strings of different lengths. As the lengths of letter strings increase, the size of the name agreement effect increases by -6.2. An interaction between length and frequency (SE = 2.343, t = -2.1, p = .037) shows that the effect of length differs for letter strings of different frequencies. As the lengths of letter strings increase the size of the frequency effect increases by -4.9. Interaction between length and AoA, as previously stated, was not significant.

Covariance parameters show a random effect of subjects on intercepts, estimated variance of 6227, and a random effect of items on intercepts, variance of 2395, in addition to the trial-level residual term.

Akaike's Information Criterion (AIC) is used to compare the goodness of fit between this model which uses name agreement as one of the predictor variables, and the second model, which uses the *H* statistic as the alternative measure of vowelisation diversity. AIC for this model using name agreement was 638005.855

6.5.4.2 Mixed effects regression model 2

The second word naming mixed effects model was run with *H* in place of name agreement and with fixed effects of participant attributes, AoA, frequency, length, and interactions between all pairs of the psycholinguistic variables. A higher value of *H* is associated with a greater number of possible alternative forms, or with alternative forms which are similar in frequency, or both. Random effects were the effect of items on intercepts and the effect of subjects on intercepts.

Table 20 Word naming model 2: Estimates of fixed effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Fixed effects							
Intercept	688.173	23.268	38.364	29.576	< .001	641.083	735.262
[GenderCode=0]	-50.413	26.862	38.000	-1.877	.068	-104.793	3.967
[GenderCode=1]	0b		.	.	.	.	.
[QualificationCode=0]	50.981	28.785	38.000	1.771	.085	-7.291	109.254
[QualificationCode=1]	0b		.	.	.	.	.
ZAge	15.443	15.001	38.000	1.030	.310	-14.924	45.811
z*H*	16.780	1.782	1428.011	9.415	< .001	13.284	20.276
zAoA	23.045	1.566	1403.274	14.716	< .001	19.973	26.117
zFrequency	-11.638	2.404	1383.389	-4.841	< .001	-16.355	-6.922
zLength	-2.011	1.781	1388.801	-1.129	.259	-5.504	1.483
zHxzL	7.597	1.792	1419.283	4.240	< .001	4.082	11.112
zHxzF	1.685	1.555	1406.633	1.083	.279	-1.365	4.735
zHxzAoA	-1.137	1.792	1411.040	-.634	.526	-4.652	2.378
zLxzF	-5.331	2.329	1376.255	-2.289	.022	-9.901	-.761
zLxzAoA	-.379	1.647	1382.778	-.230	.818	-3.611	2.852
zFxzAoA	-4.909	2.185	1406.047	-2.246	.025	-9.195	-.622
					Variance	Std. Error	
Random effects							
Item (intercept)					19662.657	126.457	
Participant (intercept)					2368.311	111.957	
Residual					6226.702	1431.972	

Notes: (1) GenderCode = 1 male 0 female (2) QualificationCode = 1 graduate 0 non-graduate (3) ZAge = age, standardised score (4) Z*H* = percentage agreement distributed across vowelisations among participants, standardised score (5) ZAoA = participant-rated mean age of acquisition by item, standardised score (6) ZFrequency = frequency of unvowelised form taken from Aralex (Boudelaa & Marslen-Wilson, 2010), standardised score (7) ZLength = abbreviation of length, standardised score (8) ZAG = abbreviation of percentage agreement among participants, standardised score (9) ZL =

abbreviation of length, standardised score (10) ZF = abbreviation of frequency, standardised score

It can be seen in Table 20 that participant attributes did not significantly affect response times: age (SE = 15.001, t = 1.0, p = .310), education (SE = 28.786, t = 1.7, p = .085), gender (SE = 26.863, t = -1.8, p = .068). As can be seen in Table 20, three of the main effects reached significance in accounting for the variance in the reaction-time data; *H* (SE = 1.782, t = 9.4, p < .001), AoA (SE = 1.566, t = 14.7, p < .001) and frequency (SE = 2.404, t = -4.8, p < .001). However, all of these were also affected by interactions with other variables and so main effects and effects of interactions are discussed together.

If all other effects are zero, then for unit increase in *H*, RT increases by 16.8ms. A significant interaction between *H* and length (SE = 1.792, t = 4.2, p < .001) shows that the effect of *H* is different for letter strings of different lengths. As the lengths of letter strings increase the size of the *H* effect increases by 7.6. Interactions between *H* and frequency (SE = 1.555, t = 1.1, p = .279) and between *H* and AoA (SE = 1.792, t = -.6, p = .596) were not significant.

If all other effects are zero, then for unit increase in AoA, RT increases by 23ms. An interaction between AoA and frequency (SE = 2.185, t = -2.246, p = .025) indicates that the effect of AoA is different for letter strings of different frequencies. As the frequencies of letter strings increase, the size of the AoA effect increases by -4.9. Interactions between AoA and *H* and between AoA and length (SE = 1.647, t = -.2, p = .818) were not significant.

If all other effects are zero, then for unit increase in frequency, RT decreases by 11.6ms. An interaction between frequency and length (SE = 1.555, t = 1, p = .279) shows that the effect of frequency differs for letter strings of different lengths. As the lengths of letter strings increase the size of the frequency effect increases by -5.3. An interaction between frequency and AoA (SE = 2.185, t = -2.2, p = .025) shows that the effect of frequency differs for words acquired at different ages. As the AoA increases, the size of the frequency effect increases by -4.9. Interaction between frequency and *H* and frequency and AoA, as previously stated, was not significant.

The effect of length alone is not significant (SE = 1.781, t = -1.1, p = .259). However, an interaction between length and *H* (SE = 1.792, t = 4.2, p < .001) shows that the effect of name agreement differs for letter strings of different lengths. As the lengths of letter strings increase, the size of the *H* effect increases by 7.6. An interaction between length and frequency (SE = 2.329, t = -2.3, p = .022) shows that the effect of length differs for letter strings of different frequencies. As the lengths of letter strings increase the size of the frequency effect increases by -5.3. Interaction between length and AoA, as previously stated, was not significant.

Covariance parameters show a random effect of subjects on intercepts, estimated variance of 6227, and a random effect of items on intercepts, variance of 2368, in addition to the trial-level residual term.

Akaike's Information Criterion (AIC) for this model using *H* is 637989.332 which is smaller than the AIC value for the first model using name agreement as one of the predictor variables. This second model is therefore to be preferred.

6.9 Discussion

This experiment investigated how native Arabic speakers name isolated words when they are presented as unvowelised letter strings, some of which are homographic and others are non-homographic. It was hypothesized that (1) response times will be shorter for non-homographic items which have only one possible pronunciation, (2) high frequency unvowelised letter strings will be named faster than low frequency unvowelised letter strings, whether they are homographic or not, (3) longer unvowelised letter strings will have longer response times, whether they are homographic or non-homographic, and (4) early-acquired unvowelised letter strings will be named faster than those acquired later, whether they are homographic or non-homographic.

Significant effects were identified with regard to (1) the number of possible alternative forms of unvowelised letter strings (as measured by the *H* statistic), (2) the age at which words were first understood (AoA), and (3) frequency as measured by frequency of unvowelised letter strings as they appeared in Aralex (Boudelaa & Marslen-Wilson, 2010). In addition, significant effects were found for interactions between *H* and length, length and frequency, and AoA and frequency. The strongest effects, taking into account interactions, were AoA, then *H*, followed by frequency. No significant interactions were found between *H* and AoA or *H* and frequency.

Investigations involving homographs in Arabic which are unvowelised letter strings cannot be compared directly with English or other Indo-European languages in which homographs are predominantly homophonic, having the same pronunciation but different spellings (Peleg & Eviatar, 2012). The process of naming homophones may be different, since phonological responses are associated with increased left brain activity, whereas visual orthography generates greater right brain activity (Peleg & Eviatar, 2007).

In the present study, participants generally took longer to name unvowelised letter strings with more alternative vowelised forms. Higher levels of name agreement were associated with shorter latencies. However, for shorter, more frequently used words with multiple vowelisations which were acquired at an early age, latencies were shorter. This indicates that naming homographs learned early in life was not problematic.

Previous studies have shown that it takes longer to make decisions about pronunciation for words with similar orthographies but different meanings, and therefore meanings may be activated before recognition (Rodd, 2004b). Where ambiguous English words have semantically related meanings, decisions are made more rapidly than if meanings are unrelated (Rodd et al., 2002). Rodd explains this in terms of attractor basins in an interactive activation account. ‘Attractors’ group together occurrences of units, in this case units of meaning, and provide stability in the representations of words. These representations also include orthographic and phonological representations and attractor basins apply equally to those representations. In the present study, higher values of *H* indicate that some of the unvowelised letter strings have more vowelised possibilities but these will generally be related in meaning through the root consonants which carry the core meaning. In Arabic, the early acquired meaning appears to be selected in word naming. Since lower name agreement is associated with longer latencies, the results of the present study suggest that other cognitive processes are involved for words acquired later.

Word naming of heterophonic homographs requires identification of a single phonological code from two or more alternatives. In English, ambiguities arising from pronunciation of irregular words can be resolved according to the Dual Route Cascaded model (Coltheart et al., 2001) by using direct connections from orthography to phonology. With unvowelised Arabic letter strings, this may be possible for high frequency or early acquired homographic and non-homographic words, but low frequency words would require access to the possible or probable forms of vowelisation through the root-and-pattern. Connectionist models such as that of Plaut et al. (1996) assume that for irregular words in English, ambiguity is resolved through the activation of nodes and weights on the connections between them, with access to meaning for low frequency words with irregular pronunciation. Arabic pronunciation is very regular once the many rules governing grammar and internal word structure have been ingrained but it cannot be directly accessed via semantic information which comes from the root. Older traditional proposed mechanisms for naming isolated words are ordered access, in which the higher frequency dominant meaning determines the choice of pronunciation (Simpson & Burgess, 1985), and exhaustive access, which proposes activation of all meanings independently of context (Onifer & Swinney, 1981). In Hebrew, it has been shown that in the absence of context, only the dominant meaning is accessed (Peleg & Eviatar, 2007). In Persian, it has been proposed that frequency determines the order of access to meanings of heterophonic homographs (Noujoumian, 2011). In the present study, although no analysis of items was undertaken other than for the correct, modal response, as many as 70% of the items deemed incorrect were possible alternative vowelisations of the homograph. This could result from individual subject exposure (participant cumulative frequency) to a particular meaning of a word, but suggests that more than the dominant meaning may be accessed and that typical measures of frequency alone do not explain why a particular pronunciation is given. The morphological family size is a possible explanation of why the correct modal responses or a closely related word was pronounced, a topic which could be investigated in future research.

The finding in the present study that longer letter strings are learned later and have longer latencies, even though they have fewer possible alternative vowelisations, suggests they are processed differently from the short, frequent, early-acquired items. Longer words may require longer and more eye fixations. Alternatively, they may require different processing as they will often have lower frequencies and so require phonological access. Saiegh-Haddad and Geva (2008) asserted that readers cannot correctly identify the correct form of an unvowelised word without accessing the word pattern as a prerequisite to phonological access. They note that in deep orthographies, individual graphemes may map onto diverse phonemes and propose that therefore larger grain size orthographic units, for example morphemes, are needed to retrieve the phonemes which make pronunciation possible. However, priming with a noun sharing a pattern of short vowels with the target word did not facilitate naming in an experiment by Abu-Rabia and Awwad (2004), which led them to propose that a noun is recognised as a whole word, and not as a combination of root and pattern, and that therefore the mental lexicon in Arabic contained at least some whole words as separate entries. Additionally. if there are many lexical entries close to each other by reason of orthographic or meaning similarities, as occurs in Arabic morphology (Boudelaa & Marslen-Wilson, 2013), access is expected to be faster.

In contrast to Abu-Rabia and Awwad, one study employing mismatch negativity (MMN) has found that Arabic roots and patterns (morphemes) may be stored separately in memory, that is, they do have specific memory traces, showing that the root produces a response in the brain earlier and in a different location than the vowel pattern (Boudelaa et al., 2010). The implications are that the orthography conveys meaning through the root, while the pattern conveys the grammatical aspects of the word, and the two together provide the phonology.

These alternatives are not necessarily mutually exclusive. In Hebrew, native speakers have been found to process Hebrew words, which are based on roots and patterns (see section 2.3.1), differently from imported words whose origin is not Semitic (Velan & Frost, 2011). Processing heterophonic homographs could involve sequential access to meaning and phonology via roots and retrieval of an appropriate short vowel pattern to complete a phonemically complete word, as in the Dual Route Cascaded model (Plaut et al., 1996). Alternatively any word learned early and encountered frequently, could be identified and pronounced via direct access to whole words in the mental lexicon, as proposed by Ibrahim and Eviatar (2012). The AoA effect is most easily explained in terms of the initial shaping of a connectionist network.

Higher word frequency is also associated with shorter latencies in a range of languages, as seen in a cross-linguistic picture-naming study involving seven languages: English, German, Italian, Spanish, Bulgarian, Hungarian and Chinese (Bates et al., 2003). The present study found a similar effect although this was outweighed by the influence of AoA and entropy (*H*). In this respect, the speed of word naming can be said to be largely determined in the present study by a combination of both earlier AoA and higher frequency (which reduce the time needed for a response) and the number of alternative forms of a homograph related to both orthography and meaning. The hypothesis that high frequency heterophonic homographs will be named faster than low frequency heterophonic homographs, on the basis that they are more likely to be stored as letter strings or complete words in the mental lexicon because of their high frequency, has therefore shown to be partially supported for the stimuli in the present study. The hypothesis for the effect of AoA is wholly supported.

The fact that later-acquired words and higher variation in percentage name agreement and number of possible alternatives increase response times has two implications for cognitive processing. The first is that searching for a single pronounceable form which has a relatively low cumulative frequency over an individual’s lifetime takes longer, and the second is that it takes longer to find one alternative among many. Both these implications suggest that known words are represented as whole words in the mental lexicon even when they are represented as letter strings in printed Arabic text, but that searching is delayed when there are few related words (possibly related in terms of root and patterns) or when there are more alternatives to choose from, which implies word patterns and vowelisations have to be identified before pronunciation. These findings appear to support the Dual Route cascaded model of word recognition (Coltheart et al., 2001). Frequently encountered non-homographic unvowelised letter strings will be stored and accessed directly from the mental lexicon. Low frequency words and multiple vowelisations of letter strings will required sub-lexical assembly. However, the distinct influence of AoA would suggest that early acquired words have a particular role in the formation and structure of the lexicon which requires additional explanation which the theory is unable to provide. Connectionist theories can only explain the influence of AoA if network simulations incorporate a developmental training element, building up vocabulary in a similar way to a child rather than training the network.

Chapter 7 LEXICAL DECISION

7.1 Introduction

This chapter investigates the effect of name agreement, Age of Acquisition (AoA), length and frequency effects on lexical decision in unvowelised printed Arabic text. It begins with hypotheses for the experiment based on the literature review, followed by background information about previous research in English, Hebrew and Arabic regarding ambiguity and homographs. Information about the participants is provided, and the materials section includes justification and explanation of how the non-word letter strings were created. A description of the procedures used is followed by statistical analysis of the data using correlation and mixed effects regression. The chapter concludes with a discussion of the findings and the implications for how unvowelised letter strings are processed in lexical decision in Arabic and for the organisation of the mental lexicon, in addition to the implications for theories of visual word recognition and reading.

7.2 Objective and hypotheses

This experiment was carried out to investigate how native Arabic speakers decide whether or not an unvowelised letter string, which may or may not be homographic, is a real word or a non-word.

Response times in lexical decision are faster for ambiguous words with many related senses than those with multiple unrelated senses in various languages, such as English (Rodd et al., 2002) and Japanese (Hino et al., 2002). The root and pattern structure of Arabic in which the root carries the core meaning (Abu-Rabia & Abu-Rahmoun, 2012) results in many unvowelised letter strings having multiple related senses as well as multiple alternative pronunciations. In a comparison of lexical decision response times to ambiguous and vowelised consonant strings in Hebrew, Bentin and Frost (1987) found that unvowelised consonant strings elicited more rapid responses than vowelized strings. This led them to propose that lexical decisions were based on the ambiguous orthography. Taouk and Coltheart (2004) also found that skilled readers recognise words in their unvowelised orthographic form rather than through their alternative phonological forms. The first hypothesis is therefore that (1) decision response latencies will be influenced by name agreement and the interaction of name agreement with other variables.

Response times in lexical decision are influenced by word frequency more strongly than in word naming (Balota & Chumbley, 1984; Balota & Spieler, 1999). Words encountered more often are associated with shorter response times (e.g. Balota & Spieler, 1999). In a series of lexical decision tasks in Hebrew, Bentin and Frost (1987) compared response times for high- and low-frequency unvowelised (ambiguous) letter strings as well as vowelized (unambiguous) strings. They found that the higher frequency alternatives elicited faster responses for both ambiguous and unambiguous letter strings. Response times were faster for ambiguous than for unambiguous letter strings, even though the ambiguous strings had both different pronunciations and different meanings. The second hypothesis is therefore that (2) decision response latencies will influenced by frequency and the interaction of frequency with other variables.

The proposal that lexical decisions in Hebrew were based on orthography led Bentin and Frost (1987) to suggest that phonological or semantic resolution of ambiguous letter strings might not be necessary. The same conclusion was reached by Share (2008) regarding Arabic. This view is supported by Taouk and Coltheart (2004) who concluded that skilled readers recognise words in their unvowelised form rather than through their phonological forms. Findings from some studies suggest that complete unvowelised shapes may represent words in the mental lexicon and be recognised as distinct entities (Abu Rabia & Awwad, 2004; Ibrahim & Eviatar, 2012). Others have proposed that skilled Arabic readers prefer to use visual-orthographic processing of unvowelised strings which has been developed over years of reading (e.g. Ibrahim, 2013). The influence of visual orthographic features in Arabic may be found in the salience of root consonants in an unvowelised letter string (Boudelaa & Marslen-Wilson, 2005, 2015) and in the early and prolonged effect of priming by roots (Boudelaa & Marslen-Wilson, 2005). The third hypothesis is therefore that (3) decisions will be influenced by the visual characteristics of the letter string.

Lexical access is affected by many variables including orthographic similarity and word frequency (e.g., Cutler & Davis, 2012), orthographic depth and AoA (Balota et al., 2004; Adelman, 2012). A review by Juhasz (2005) found that the AoA effect had been identified in word naming, lexical decision and picture naming. Brysbaert & Ghyselinck (2006) proposed two types of AoA effect, one dependent on frequency and the other not; the frequency-independent effect was hypothesized to result from competition at the level of meaning. Support for the role of semantics in explaining lexical decision has been provided by Cortese & Khanna (2007) who found a larger effect of AoA in lexical decision than in word naming tasks. AoA has been found to exert a strong effect on response times in various languages including English (Kuperman et al., 2011) and Spanish (González-Nosti et al., 2014). The fourth hypothesis is therefore that (4) decision response latencies will be influenced by AoA and the interaction of AoA with other variables.

In lexical decision in English, the effect of word length has been shown to differ according to the length of the word (New et al., 2006). Length had no effect on response times for words of five to eight letters, although words of three to five letters elicited faster responses, and words of eight or more letters were responded to more slowly. Evidence relevant to the present study from Semitic languages is sparse. One study in Hebrew found that lexical decision response times to six-letter unpointed strings were significantly slower than response times to three-letter strings (Lavidor & Whitney, 2005). A later study in Hebrew found that longer unpointed letter strings were read more accurately and rapidly than longer, fully pointed words by skilled Hebrew readers. They also found that typical readers responded more quickly to longer unpointed words than to shorter ones (Weiss et al., 2015). The fifth hypothesis is therefore that (5) response times will be slower for longer words.

7.3 Background

Lexical decision tasks are considered to give more direct access to the lexicon. There has been a general assumption that all variants of the task involve matching features of the stimuli to entries in the mental lexicon, and that frequency plays a key role in decisions (Balota & Chumbley, 1984). Chumbley and Balota (1984) also proposed there were two stages involved rather than single-stage direct access to the lexicon. Difficulty is said to be measured more directly in lexical decision tasks because they avoid the complication of reading aloud. In English, word length can affect results although the effect seems stronger in low frequency words (Balota et al., 2004) thus word length should be matched in pairs of real words and non-words.

Lexical decision tasks haves been used in Arabic studies to investigate word recognition in various ways. As discussed in the literature review in chapter three, these include the impact of diglossia (Bentin & Ibrahim, 1996), the validity of the etymon as a construct (Mahfoudi, 2007), and the impact of connectivity (Khateb et al., 2013) among others. Lexical decision has also been used to investigate the effect of homographs on selection of low- or high-frequency alternatives, and the effect of vowelisation on the speed of decision (Roman, & Pavard, 1987, and Bentin & Frost, 1987, respectively).

Lexical decision experiments in Hebrew have led researchers to propose that ambiguous letter strings do not need to be resolved phonologically or semantically (e.g., Bentin & Frost, 1987) while Share (2008) drew the same conclusion regarding Arabic. According to Share, resolving the ambiguity depends on a combination of vocabulary knowledge and application of the rules of syntax; what words does someone know that share the root consonants and what short vowel patterns can be inserted. Differences in processing Arabic, English and Hebrew orthographies have been identified (Ibrahim & Eviatar, 2012). It has been shown that in young children, ability to remember letter strings in Hebrew played an important role in learning to read (Shatil & Share, 2003), leading Assad and Eviatar (2014) to propose that visual factors play a larger role in reading in Semitic languages than in English and other Indo-European languages because the root consonants stand out clearly in an unvowelised letter string and fully vowelised Semitic languages are more cognitively demanding.

Different kinds of ambiguity require different kinds of resolution. The extent of ambiguity in Arabic, as shown by participants in the name agreement study presented in chapter 4, indicated that a single ambiguous letter string may be ambiguous in two ways. There may be ambiguity in the mappings between unvowelised spelling and alternate vowelisations, hence between spelling and pronunciation and thus meaning. Some letter strings which have shared unvowelised spelling can have the same pronunciation but have unrelated meanings. In English, longer response times to homophones with similar spellings were reported by Haigh and Jared (2007), leading to proposals that access to phonological representations was necessary before an orthographic representation could be identified in order to decide whether a stimulus was a real word. Rodd et al. (2002) distinguished between words with many related senses and those with unrelated meanings, proposing that the former enhanced decisions while the latter hindered them. In such situations, the mapping of letter string to output may be arbitrary. Cortese and Khanna (2007) proposed that effects of AoA are likely to be stronger in situations where mapping is arbitrary. They also proposed that this implied an AoA effect was stronger in lexical decision when compared to word naming.

Researchers face challenges in designing lexical decision tasks in Arabic for three reasons, related to finding non-words that are phonologically and orthographically legal (pseudowords) in printed Arabic. Firstly, Arabic is written without the short vowels for adult readers so that makes the possibility of finding non-words more difficult than in languages such as English where all the vowels are part of the printed words and can be exchanged or replaced more easily. Secondly, changing one of the consonants in the root frequently results in another root, generating a different range of letter strings for real words. Finally, the rules governing the internal structures of Arabic words further limit the availability of alternatives. It was important to use non-word letter strings which closely resembled real word letter strings in order to ensure that homography remained the focus of the experiment. Inverting a letter or exchanging letters as in many studies would have tested recognition of letter characteristics as well as, or perhaps instead of, the effect of homography.

Lexical decision latencies have been shown to be affected by several psycholinguistic variables. More frequently encountered words are associated with faster decisions (e.g., Balota & Spieler, 1999), as are words encountered earlier in life (Cortese & Khanna, 2007). In English, the consistency of orthography has been shown to influence response times, inconsistent spellings leading to longer latencies (Balota et al., 2006). In English, word length has been found to have a different effect depending on the length of the word (New et al., 2006). Decisions were faster for words of three to five letters, and slower for words with 8 to 13 letters, but response times for words of 5 to 8 letters were unaffected by length.

7.4 Method

7.4.1 Participants

A total of 40 adult native speakers of Arabic (20 males and 20 females, mean age 29.15 years, standard deviation 9.802, age range: 18-50 years) were recruited from a sample comprising 20 individuals from the population of Saudi international students and their spouses in the UK and a further 20 adults in Riyadh, Saudi Arabia. 21 participants were educated to bachelor degree level or above, while the remaining 19 had completed their education before obtaining a bachelor degree. The mean number of hours spent reading MSA per day was 2.13 hours (SD of 1.04, range: 0.5 to 5 hours). Participants received 15 pounds for travel expenses. All were native speakers and had normal or corrected-to-normal vision. None reported any language problems.

Ethical approval was obtained from Lancaster University as well as Imam Muhammad University in Riyadh, Saudi Arabia (see Appendix C).

7.4.2 Materials

For the stimuli, 1,352 letter strings representing real words were taken from the database of 1,474 described in Chapter 4. Of the 1,352 unvowelised letter strings, 794 (58.7%) are homographs (single unvowelised letter string with more than one possible pronunciation and meaning), and the remaining 558 have a unique phonological and semantic form. A list of 1,352 non-word letter strings which pronounceable non-words matched for length with the letter strings representing real words.

Constraints of the internal structure of Arabic words imposed by the existence of more than 5,000 consonantal roots made it impossible to create more than 1,352 legal unvowelised non-word letter strings. For letter strings representing real words, frequency estimates, expressed as frequency per million, for the unvowelised strings of letters, as they appeared in printed Arabic text, were taken from Aralex, a corpus of 40 million MSA words drawn from modern sources (Boudelaa & Marslen-Wilson, 2010). AoA ratings for the modal vowelised forms of real words were provided by a survey of 89 participants, as detailed in Chapter 5. For words with more than one alternate vowelisation, according to the name agreement survey, the modal form was the form most frequently used by participants in response to the unvowelised letter strings.

Non-words can be unpronounceable due to combinations of letters which are not found in the language, 'zpks' in English for example, and orthographically illegal, for example by including letters from a different alphabet or by inverting letters. They can also be phonologically and orthographically legal. If non-words are unpronounceable it will disturb the results because they will be processed differently, so that attention may given to trying out silently how they might be pronounced rather than focusing on the word or non-word decision (Yap et al., 2006). Using the sublexical route in a DRC model (Coltheart et al., 2001) would result in a dead end, while there would be no connection weights or patterns of activity in a connectionist model (Plaut et al., 1996).

The procedure of non-word selection used for this study was as follows. Each non-word created was matched for length against an unvowelised string which was a real word. One of the three consonants of the root was substituted for another to produce words that were orthographically and phonologically legal but lacking meaning. If this produced another real root, a second consonant was substituted. If it was not possible to produce a false root by substituting one or two consonants, the suffix was changed. As far as possible, the overall shape of the word was maintained in order to comply with the rules regarding Arabic characters. For example, if a real word letter string contained two occurrences of the same letter, this feature was retained in the non-word letter string. The final characters and any long vowels present in the letter string representing a real word were retained in the non-word letter string. Characters occur in groups and while some can be attached to adjacent letters, others cannot (Saiegh-Haddad & Henkin-Roitfarb, 2014). However, disambiguation of words in Modern Standard Arabic can be problematic in the absence of diacritics (Boudelaa & Marslen-Wilson 2001), therefore when a non-word letter string was created from a real word letter string, groups of letters which differed only in terms of the number of dots were avoided in order not to mislead the study participants. The list of non-words was then checked by three Arabic language experts to ensure it contained no unvowelised letter strings which were real members of the MSA vocabulary.

Certain variables have been shown to exert considerable influence in lexical decision. The first of these is word frequency, which is especially important in lexical decision tasks, in which it can account for more than 30% of the variance in response times (Balota et al., 2004; Brysbaert & New, 2009). The second is word length, which is measured in characters or syllables and sometimes in both, and has been shown to be influential in both word naming and lexical decision tasks (Ferrand et al., 2011). The third variable refers to the similarity of one word to other words, or neighbourhood density. In English, this has typically been measured in terms of the average frequency of the letter pairs in the word, or alternatively in terms of the orthographic neighbourhood which has been defined as the number of words that can be formed by changing a single letter in the target stimulus (Balota et al., 2000). The appropriateness of measures of neighbourhood density in Semitic languages is examined in section 2.4.3. The fourth important variable, in word naming, is held to be the first phoneme of a word because of its influence on of monosyllabic words, although the first phoneme has little effect in lexical decision tasks (Balota et al., 2004).

In order to give an idea of the relative importance of variables, when these four variables were subjected to stepwise multiple regression analysis by Yarkoni, Balota and Yap (2008), the most influential variable in prediction of lexical decision time was word frequency, accounting for just over 40% of the variance. The second most influential variable was a measure of neighbourhood density, which accounted for almost 13%; the measure used was OLD20 which is calculated as the mean Levenshtein Distance (LD) from a word to the twenty words with the most similar orthography based on the fewest number of additions, deletions, and substitutions of letters (Yarkoni et al, 2008). In a different study, word length measured in syllables accounted for just over 1% in lexical decision, while the combined effect of other variables accounted for 2.0% (Brysbaert et al., 2011). The measured impact of variables is not the same in all languages, or indeed in all experiments depending on their design and the method of analysis employed. The following variables are important for this study, hence the decision to include them as specific item characteristics of the unvowelised letter strings representing real words: frequency, word length and AoA. Table 21 summarises item characteristics, providing the mean, SD, minimum, and maximum values for each variable.

Table 21 Summary of item characteristics (lexical decision) (N = 1,352)

Variable	Mean	SD	Min.	Max.	Range
Agreement	.773	.234	.210	1.000	.790
H	.736	.756	.001	2.94	2.94
Type	2.420	1.744	1	11	10
MeanAoA	10.695	2.807	3.800	19.100	15.3
Frequency	57.485	133.663	.010	2065.76	2065.75
Length	4.250	.985	2	8	6

Distributions of frequency, word length and AoA are presented in Figures 18 to 20.

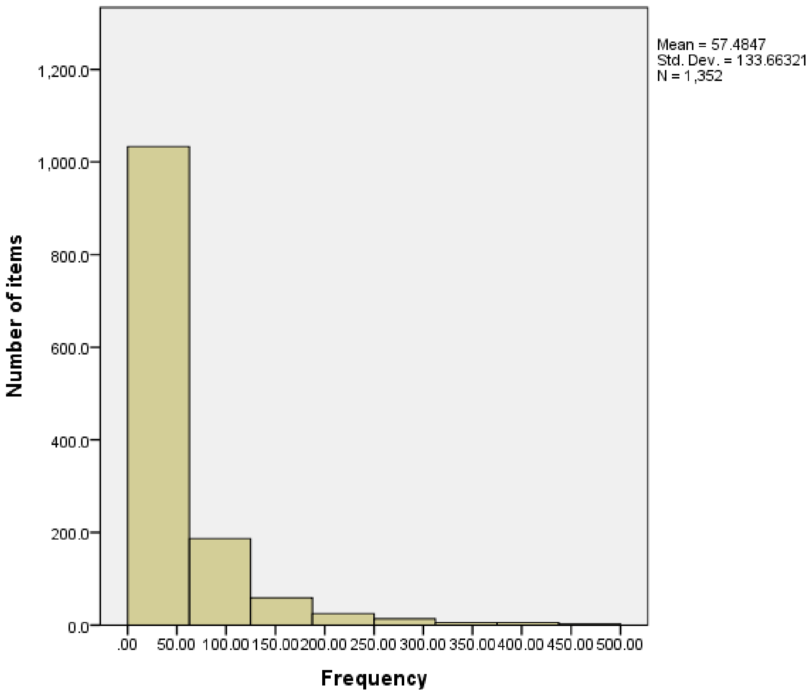


Figure 18 Distribution of frequencies for real word strings in lexical decision

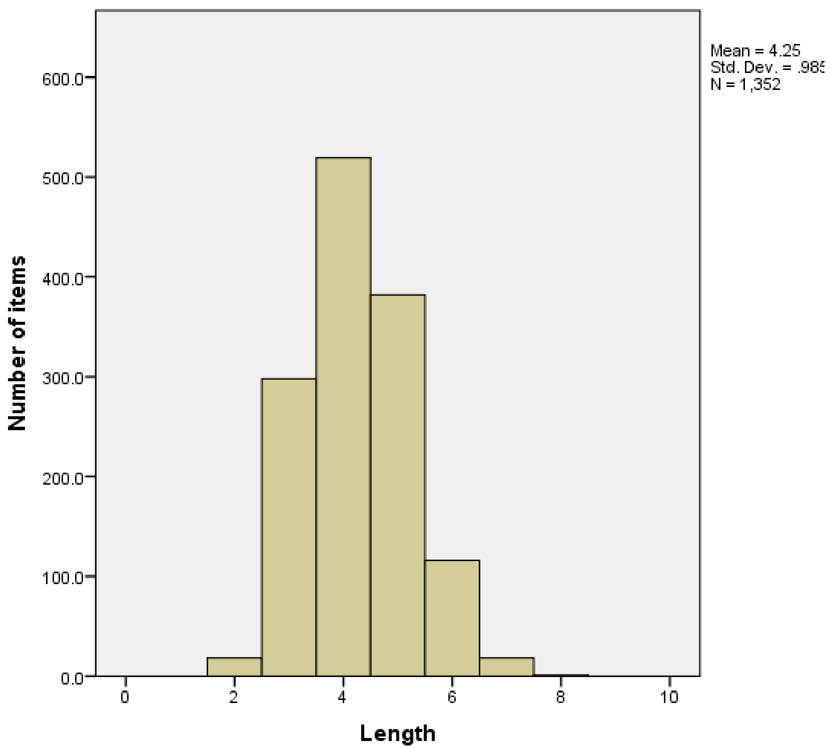


Figure 19 Distribution of word length for real word strings in lexical decision

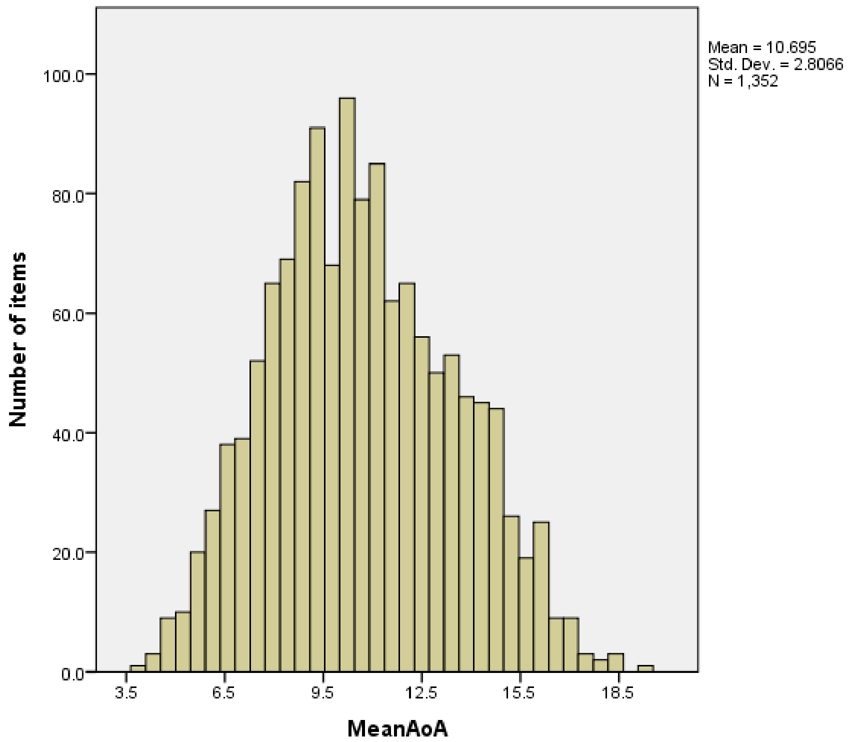


Figure 20 Distribution of AoA for real word strings in lexical decision

Information sheets were distributed among Saudi students and their dependants who were studying in the UK. The information sheet contains information about the purpose of the experiments, the incentive offered, the right to withdraw and the methods to be used if the participant wished to take part. When the participant agreed to take part, they had to sign a consent form, after which they were asked to give some basic personal information (age, gender, education level, the number of hours spent reading MSA per day).

7.4.3 Apparatus and procedure

Subjects were presented with 1,352 unvowelised letter strings corresponding to real word vowelisations and 1,352 unvowelised letter strings that did not correspond to real word vowelisations. They were instructed to decide as rapidly and accurately as possible whether each letter string was a real word. The DMDX application (Forster &

Forster, 2003) was used to present stimuli and to record responses on a Windows XP laptop. Responses were indicated by keypress, recorded using the laptop keyboard. Participants were asked to press the L key if they recognised the stimulus as a word. They were asked to press the A key if they recognised the stimulus as corresponding to a non-word.

Word and non-word stimuli were divided into 19 blocks of 141, 142 or 143 items. The order of blocks, and of trials within blocks, was randomized for each participant. Participants were allowed to take breaks between trial blocks. The complete task took approximately 3 hours to complete, on average.

Participants were seen individually in a quiet room, where the researcher instructed them to look at each letter string and decide as quickly and accurately as possible, whether or not the letter string represented a real word. Instructions appeared on the screen of the computer where they were seated prior to the start of the task and were read aloud by the experimenter. Participants had to complete the task in a single session.

For each trial, each participant was seated 50 cm from a computer screen. A fixation point was presented at the centre of the screen for 500 ms, followed straightaway by presentation of the stimulus in black size 20 Arial font at the centre of the screen where it remained for 2,500 ms. A white screen then appeared for 500ms.

Participants completed 20 practice trials consisting of 20 unvowelised letter strings representing real words and 20 unvowelised non-word letter strings, after which they were given the opportunity to ask any questions.

7.5 Results

7.5.1 Data preparation

A total of 108,160 responses were recorded for the task, 54,080 responses to letter strings corresponding to real word vowelisations, and 54,080 responses to letter strings that did not correspond to real word vowelisations. Only the latencies of correct responses to real words were analysed, excluding 5,081 (9.39%) responses (errors or null responses) to leave 48,999 observations for analysis.

7.5.2 Data analysis

Mean response times were calculated for all real-word items (M = 809.69, SD = 141.35, range: 550 to 1765 ms). Distribution of mean response times for all items is shown in Figure 21.

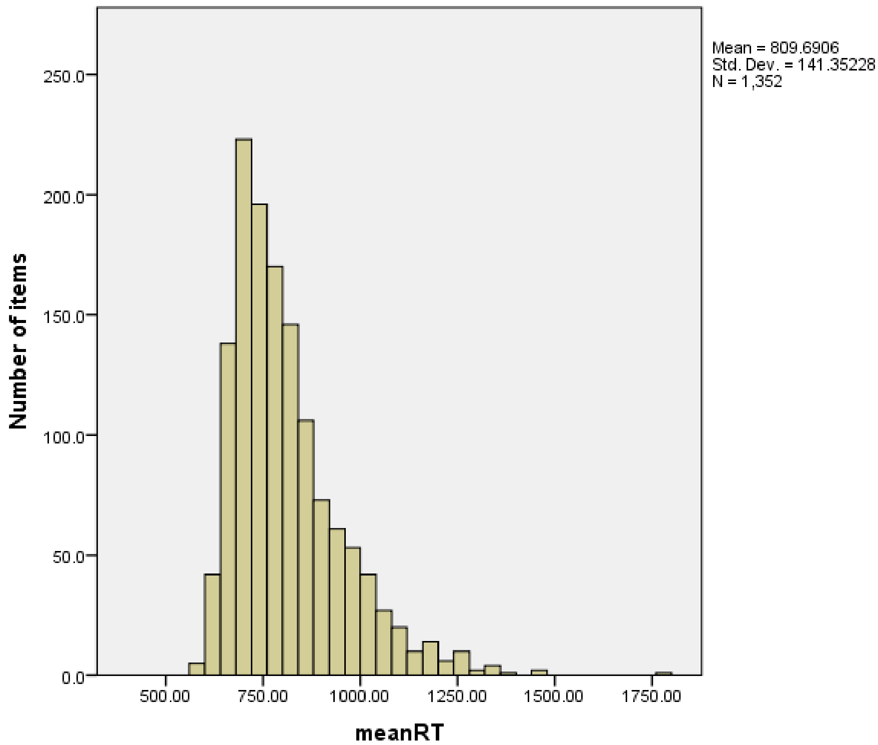


Figure 21 Distribution of mean response times for all items: lexical decision

Correlation and mixed effects regression were used to analyse data. Mixed effects regression using IBM SPSS (version 22) was used for analysis for several reasons. Analysis of continuous variables such as the *H* statistic provides a more realistic picture of what is happening that arbitrarily assigning vowelisation diversity to ‘high’ or ‘low’. The effects of interactions between predictor variables such as *H* and AoA can be investigated in addition to the effects of individual predictor variables and can increase understanding of their respective contributions to response times. made by the independent variables to the outcome variable, without overemphasizing or underplaying the influence of one or more variables. The inclusion of random effects of subjects on intercepts and random effects of items on intercepts enables non-repeatable variables, such as random differences between participants, to be taken into account.

Raw response times were used in the mixed effects modelling. However, mean response times per item were calculated in order to establish correlations between variables.

7.5.3 Correlations

Correlations run using IBM SPSS (version 22) are presented in Table 22. Correlations were calculated to identify which variables had a significant relationship with other variables. This indicates how the effects relate to the response times and to each other. Correlations were also used to identify possible instances of collinearity.

Table 22 Summary of Pearson's r bivariate correlations between key psycholinguistic variables (N = 1,352 in all comparisons)

		H	Agree-ment	Type	AoA	Frequency	Length
MeanRT	r	-.075**	.077**	-.086**	.664**	-.253**	.130**
	p	.006	.004	.002	< .001	< .001	< .001
H	r		-.968**	.931**	-.109**	.087**	-.513**
	p		< .001	< .001	< .001	.001	< .001
Agreement	r			-.839**	.092**	-.081**	.490**
	p			< .001	.001	.003	< .001
Type	r				-.128**	.097**	-.510**
	p				< .001	< .001	< .001
AoA	r					-.235**	.256**
	p					< .001	< .001
frequency	r						-.137**
	p						< .001

Notes: (1) Mean RT = mean response time. (2) *H* = *H* statistic (distribution of agreement across all correctly diacritised forms) (3) Agreement = percentage name agreement among participants (4) Type = number of different alternative vowelisations of homographic letter strings (1 to 11) (5) AoA = participant-rated mean age of acquisition by item (6) frequency from Aralex (Boudelaa & Marslen-Wilson, 2010) (7) length = length of word in letters

There was a weak negative correlation between mean RT and the *H* statistic, longer response times corresponding to a lower value of *H* (r = -.075, p = .006), a weak positive correlation between mean RT and agreement, longer RTs being associated with higher name agreement (r = .077, p = .004) and a weak negative correlation between mean RT and type, with shorter response times to letter strings with more alternative vowelisations (r = -.086, p = .002). There was a strong positive correlation between mean RT and AoA, a later AoA being associated with a longer RT ($r = .664, p < .001$), a weak negative correlation between mean RT and frequency, higher frequency letter strings elicited shorter RTs ($r = -.253, p < .001$) and a weak positive correlation between mean RT and length, longer RT corresponding to longer word length ($r = .130, p < .001$).

There was a strong negative correlation between *H* and agreement, lower values of *H* corresponded to higher levels of name agreement ($r = -.968, p < .001$), a strong positive correlation between *H* and type, with higher values of *H* associated with a greater number of different vowelisation types ($r = .931, p < .001$), a weak negative correlation between H and AoA, indicating that higher values of H corresponded to earlier AoA ($r = -.109, p < .001$), a weak positive correlation between H and frequency, such that higher values of H were associated with higher frequency words ($r = .087, p < .001$), and a moderate negative correlation between H and length, such that lower agreement as measured by H was associated with increased word length ($r = -.513, p < .001$). For agreement, there was a strong negative correlation with type, lower agreement associated with more possible types ($r = -.839, p < .001$), a weak positive correlation with AoA, with lower agreement corresponding to later AoA ($r = .092, p = .001$), a weak negative correlation with frequency, such that lower agreement was associated with higher frequency ($r = .837, p < .003$), and a moderate positive correlation with length, such that higher agreement was associated with greater length ($r = .490, p < .001$).

For type, there was a weak negative correlation with AoA ($r = -.128, p < .001$), letter strings with more alternative vowelisations being associated by earlier AoA, a weak positive correlation with frequency ($r = .097, p < .001$), higher frequency words associated with more alternative vowelisations, and a moderate negative correlation with length ($r = -.510$, $p < .001$), longer words associated with fewer alternative vowelisations. There was a weak negative correlation between AoA and frequency ($r = -.235$, $p < .001$) more frequently encountered words associated with learning at an earlier age, and a weak positive correlation between AoA and length ($r = .256$, $p < .001$), with longer words being acquired later. Finally, there was a weak negative correlation between frequency and length ($r = -.137$, $p < .001$), higher frequency words associated with fewer letters in a letter string.

7.5.4 Mixed effects regression

A mixed effects model was tested in order to examine the influence of the predictor variables on responses times. Fixed effects were specified as: vowelisation diversity using *H* (this variable having been shown in the name agreement study and word naming experiment to be a fixed effect in the model which better fitted the data); letter string frequency from Aralex (Boudelaa & Marslen-Wilson, 2010); word length in letters; and rated AoA. Fixed effects were also specified for participant attributes of age, gender and level of education. Finally, we specified interaction terms due to the effects of interactions between vowelisation diversity (*H*) and length or frequency or AoA, between length and frequency or AoA, and between frequency and AoA. As before, continuous numeric variables were standardized before entry as predictors in models. Interaction terms were formed as multiplicative products of lower-order component variables. Random effects were specified as random effects of subjects on intercepts and random effects of items on intercepts. The inclusion of these random effects terms allowed us to take into account unexplained differences between participants or between items in the overall response latency. A summary of effects estimates is presented in Table 23.

Table 23 Lexical decision model: Estimates of fixed effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Fixed effects							
Intercept	753.905	47.298	54.860	15.939	< .001	659.113	848.698
[GenderCode=0]	83.127	41.867	119.633	1.985	.049	.229	166.024
[GenderCode=1]	.000		.	.	.	.	.
[QualificationCode=0]	6.022	58.326	41.113	.103	.918	-111.759	123.803
[QualificationCode=1]	.000		.	.	.	.	.
zAge	-1.427	30.867	44.719	-.046	.963	-63.606	60.753
z*H*	-1.057	3.460	1242.731	-.305	.760	-7.844	5.731
zAoA	81.289	2.924	1290.855	27.798	< .001	75.552	87.026
zFrequency	-38.413	4.397	1205.754	-8.736	< .001	-47.039	-29.787
zLength	-9.901	3.364	1244.916	-2.943	.003	-16.500	-3.301
zHxzL	5.240	3.318	1259.538	1.579	.115	-1.269	11.749
zHxzF	-1.784	2.862	1206.717	-.623	.533	-7.399	3.831
zHxzAoA	1.725	3.340	1315.952	.516	.606	-4.827	8.276
zLxzF	-11.094	4.924	1207.732	-2.253	.024	-20.755	-1.433
zLxzAoA	8.926	2.993	1301.616	2.982	.003	3.054	14.798
zFxzAoA	-27.776	4.033	1214.751	-6.887	< .001	-35.688	-19.864
Random effects							
				Variance	Std. Error		
Item (intercept)				22969	5228.14		
Subject (intercept)				7289	386.796		
Residual				85037	551.742		

Notes: (1) ZAoA = participant-rated mean age of acquisition by item, standardised score (2) ZFrequency = frequency from Aralex (Boudelaa & Marslen-Wilson, 2010), standardised score (3) ZLength = length of letter strings in letters, standardised score (4) ZL = abbreviation of length, standardised score (5) ZF = abbreviation of frequency, standardised score (6) ZH = standardised score for H statistic, measure of distribution of percentage agreement among alternative vowelisations

If all other effects are zero, then for unit increase in AoA, RT increases by 81.3ms. A significant interaction between AoA and length (SE = 2.993, t = 3.0, p = .003) shows that the effect of AoA is different for letter strings of different lengths. As the lengths of letter strings increase the size of the AoA effect increases by 8.9. The interaction between AoA and frequency was also significant (SE = 4.033, t = -6.9, p < .001). As frequency of the unvowelised letter string increases, the size of the AoA effect increases by -27.8. Interaction between *H* and AoA was not significant (SE = 3.340, t = .5, p = .606).

If all other effects are zero, then for unit increase in frequency, RT decreases by 38.4ms. An interaction between AoA and frequency (SE = 4.033, t = -6.9, p < .001) indicates that the effect of frequency is different for different AoA. As AoA decreases, the size of the frequency effect increases by -27.8. An interaction between frequency and length (SE = 4.924, t = -2.3, p = .024) indicates that the effect of frequency differs for different lengths of letter strings. As the lengths of letter strings increase, the size of the frequency effect increases by -11.094. The interaction of frequency and *H* was not significant (SE = 2.862, t = -.6, p = .533).

If all other effects are zero, then for unit increase in length, RT decreases by 9.9ms. The interaction between length and frequency has been previously mentioned. An interaction between length and AoA (SE = 2.993, t = 3, p = .003) shows that the effect of length differs for different AoAs. As the AoA increases, the size of the length effect increases by 8.9. Interaction between length and *H* was not significant (SE = 3.318, t = 1.6, p = .115).

Covariance parameters show a random effect of subjects on intercepts, estimated variance of 7289, and a random effect of items on intercepts, variance of 22969, in addition to the trial-level residual term.

7.5 Discussion

It was hypothesized that (1) decision response latencies will be influenced by name agreement and the interaction of name agreement with other variables, (2) decision response latencies will be influenced by frequency and the interaction of frequency with other variables (3) decisions will be influenced by visual characteristics of the letter string decisions, that (4) decision response latencies will be influenced by AoA and the interaction of AoA with other variables, and that (5) response times will be slower for longer words.

Results showed that name agreement, as measured by *H*, had no significant effect on latencies, neither on its own nor in interactions with the other variables. This is somewhat surprising since the literature posits that there will be a difference, for example in Hebrew (e.g., Bentin & Frost, 1987) and English (e.g., Rodd et al., 2004). The first hypothesis was therefore not supported. The same finding strongly suggests the same cognitive processes are at work for homographs and non-homographs. It also suggests that lexical decisions are based on surface features of the unvowelised letter string, whether the complete string, or the three root consonants.

Frequency has a strong effect on response times, an effect which is stronger when frequency interacts with AoA. The interaction of frequency with length also has the effect of increasing the size of the frequency effect. The second hypothesis is therefore supported. The absence of an effect of H also indicates that disambiguation is not needed and therefore phonological access is not required. Instead the decision appears to be based on visual characteristics. Hence the third hypothesis is supported.

The effect of AoA was strong and significant, the strongest single predictor variable in this experiment. Allowing for the fact that few variables were included, and that the contribution of AoA might have been less if it had been one of twenty variables, there is strong evidence for the independence of AoA and frequency, as well as a combined effect. This is consistent with the explanation offered by Brysbaert and Ghyselinck (2006) that there were two kinds of AoA. One kind of AoA is related to learning that is reinforced over time by usage, the frequency-dependent effect, and the other is related to early links made between meaning and pronunciation of a word and is independent of frequency, the semantic effect. Brysbaert and Ghyselinck (2006) note that the AoA effect shows up as clearly distinct from the effect of frequency on picture naming. It is widely reported that words learned early in life can be accessed and pronounced more rapidly than those acquired later (e.g., Johnston & Barry, 2006). In their reanalysis of 30,000 words taken from the English lexicon project, Kuperman et al. (2011) showed that AoA had strongly affected standardised response times, second only to log frequency. In Spanish, lexical decision results have indicated a significant effect of AoA and frequency irrespective of the kinds of stimuli employed (González-Nosti et al., 2014).

AoA effects vary from language to language, according to the transparency of the orthography. They are seen to emerge in particular where the mapping between input and output is inconsistent (Cortese & Khanna 2007). Although the mapping between input and output can be considered consistent in Arabic, it may be that due to the complex rules involved, AoA has a stronger influence where rules are complex and learned in the early years of school. Certainly the effect of AoA increases with the length of unvowelised letter string. The fourth hypothesis is therefore supported.

Length by itself was facilitatory. One possible explanation is that longer words typically had fewer orthographical, phonological, semantic or morphological neighbours, and were less likely to have multiple alternative vowelisations. This is supported by the finding for word naming that as length increased, although longer words had fewer alternatives, they took longer to name. The effect of the interaction between length and AoA meant that the effect of length was increased as AoA increased. However, the finding from the re-analysis of data from the English Lexicon Project that whether word length was inhibitory or facilitatory depended on word length (New et al., 2006) was not investigated for Arabic in the present study. Longer words learned later could be attributed to their lengths needing multiple eye fixations. The fifth hypothesis is therefore not supported since the faciltatory effect of length alone is offset by the interaction of frequency and AoA.

The facilitatory effect of frequency suggests that more frequently encountered words may be processed differently from longer and lower frequency words. This is consistent with earlier research (e.g. Balota et al., 2004). The frequency effect is explained in the DRC model (Coltheart et al., 2001) by access through the lexical route. Comectionist models (e.g. Plaut et al., 1996) explain it by the repeated exposure to a word making the connections in the network stronger.

Bentin and Frost (1987) found that, in Hebrew, lexical decision response times for heterophonic homographs were faster than those for pointed and therefore unambiguous alternatives. They argue that this implied that lexical access was based on the printed text that represented the various phonological alternatives without recourse to the phonology itself (Frost, 2003). In the absence of any significant effect of variables measuring the degree of homography in the present experiment, the same inference is drawn. Evidence from research into the functioning of the brain indicates the something different is happening in language processing in the Arabic brain compared with the English brain. A functional magnetic resonance imaging (fMRI) investigation into the processing of Arabic diacritics indicated that unvowelised words increased activation in a different brain area from the area activated by words with diacritics, leading the researchers to propose that unvowelised words led to greater activity in searching lexical entries (A.K. Bourisly, Haynes, N. Bourisly, & Mody, 2013). Similar findings were reported for an fMRI study in Hebrew (Weiss et al., 2015). Results of the present study regarding length may point towards different processes being involved in lexical decision. The effect of length alone is that longer words have shorter latencies, perhaps because the additional letters reduce the number of neighbours or constrain the number of available alternatives in heterophonic homographs, leading to faster decision. The effect of length x AoA is that decisions for longer words learned later take more time. These processes can be explained by the DRC model of reading, with the faster lexical route for words learned early and the sublexical route for longer words learned later the latter.

No significant effect of *H* was found. There was no significant difference in RTs between ambiguous and unambiguous words, suggesting that the same processes are at work. Decisions are made for both homographic and non-homographic letter strings on the basis of what is seen in the string. It is proposed here that the decision is probably based on the correct appearance of a recognised root, because roots carry meaning. This provides support for the third hypothesis.

Decisions were faster for high frequency unvowelised letter strings, as expected. The second hypothesis is therefore supported. Decisions on the basis on the visual information in the letter string mean that phonological access was not used, since the word naming experiment in the present study showed significant differences in RTs when access to phonemes was required for pronunciation. The third hypothesis is therefore also supported.

Early acquired words were decided on more quickly and the fourth hypothesis is supported. However, the strength of the AoA effect is surprising, even though few variables were controlled for. The implication of the strength of the AoA effect confirms the need to include the way in which language is acquired in models, as suggested by Frost (2012).

The effect of word length alone was interesting; longer letter strings elicited faster response times. Taking into account the interactions of length with frequency and AoA, it is not possible to state that response times will be slower for longer words.

The absence of an effect of name agreement in lexical decision, in contrast to its significant effect on its own and in its interaction with length in word naming, suggests that phonological disambiguation was not involved in all decisions. However, decisions for infrequently encountered words may require lexical access to individual phonemes via the root. The smaller effect of name agreement nonetheless points to lexical decision for high frequency and earlier acquired words being based on visual cues (the root consonants or possibly the unvowelised letter string) suggests that lexical access may not have been involved. The Dual Route Cascaded model (Coltheart et al., 2001) can explain this provided the element of interactive-activation is incorporated in the model, with direct output from the salient orthographic features. In addition, a morphological lexicon containing roots would be required for recognition.

A developmental connectionist model, for instance Monaghan and Ellis (2010) would, as indicated in the word naming chapter (Chapter 6), could account for the effects on the basis of the differing weights on the connections and simultaneous processing across the network, which would need to be trained in stages to replicate the processes of learning to read as closely as possible.

Chapter 8 GENERAL DISCUSSION

8.1 Introduction

This research aimed to shed light on how reading in Arabic is affected by a highly homographic orthography. In order to achieve this, word naming and lexical decision were selected as the tasks to examine if some of the key variables predicting RTs were similar or different in Arabic as compared with Hebrew, another Semitic language, and with English and other languages where relevant. This chapter starts by reviewing the two studies and the main findings of the two experiments, after which consideration is given to whether the findings answer the research questions. It examines whether the findings are adequately explained by existing theoretical models of visual word recognition and reading aloud. The chapter concludes with the limitations of the present study and suggestions for future research.

8.2 The studies: name agreement and AoA

It was necessary to create a database of 1,474 unvowelised letter strings, which included unvowelised letter strings which had only one spelling and meaning and others which had more than one spelling and meaning (887 heterophonic homographs and 587 non-homographic words). Actual frequencies for selected letter strings were obtained from Aralex (Boudelaa & Marslen-Wilson, 2010) and word length confirmed. A name agreement study was conducted with 445 participants in order to identify which words were homographs and to determine how many forms each had, also which form should be considered the dominant meaning as identified by the highest proportion of participants giving this as the first form and meaning they thought of. Homographic unvowelised letter strings were found to have from 2 to 11 types (alternative vowelisations) as recorded by participants, with a mean of

3.4. It is interesting that 60% of the letter strings used in the present research were identified as homographs because it was not known at the start of the study how many letter strings participants would understand as homographs. The remaining unvowelised letter strings elicited a single vowelisation and pronunciation, indicating that the problems associated with reading homographic text may have been overstated in the past.

A further study was conducted with 89 different participants to provide subjective AoA ratings for each of the agreed vowelised forms of the 1,474 words. In addition to the database, a list of 1,352 legal non-words was created for the purpose of conducting future experiments. It proved challenging to create non-words which were matched for length and complied with the morphological structure of Arabic words without creating an existing real word due to the constraints of roots and patterns. However, this method was preferred to straightforward letter transposition as typically used in English and European languages because native speakers of Hebrew had been shown to be particularly sensitive to such transposition (Velan & Frost, 2007). Lexical space was considered to be organised by morphological families based on roots rather than by a simpler orthographical arrangement (Velan & Frost, 2007). One explanation for the strong effect of AoA is that early-acquired words in Modern Standard Arabic (MSA) are learned many times over with plenty of repetition in reading aloud in the classroom between the ages of six and twelve; first as sounds, then as fully vowelised letter strings, then as partially vowelised letter strings, and finally without diacritics.

8.3 Word naming

The database was used in a word naming study with 38 native speakers of Arabic. Word naming of unvowelised Arabic words places a heavier perceptual and cognitive load on readers than English or Hebrew, even though they are all considered to be relatively deep orthographies. The perceptual load in unvowelised text relates largely to the cursive nature of printed text and the lack of regular spacing of letters. The cognitive load refers to the requirements for additional processing of morphemes in word naming or reading aloud. The absence of short vowels in printed text allows the consonantal root to stand out. The root conveys core meaning and at the same time requires readers to supply the missing letters in order to be able to read words aloud, since some essential phonological components of the word are not present (Frost, 2012; Ibrahim et al., 2013). Complexities in Arabic which are not found in Hebrew include not simply cursive letters but the fact that they are written differently depending on their position in the word. Each letter can be pronounced in three different ways according to the diacritic (short vowel) attached to it which in turn depends on the word. Root letters are intertwined with patterns of vowels, affixes, infixes and suffixes, and words are inflected, resulting in the additional cognitive load required to correctly identify and understand what is read (e.g., Eviatar, Ibrahim, & Ganayim, 2004).

It has often been asserted that Arabic homographs cannot be read correctly if they are presented as isolated words without context, and few studies in either Hebrew or Arabic have focused on more than 50 to 100 such words in lists. Thus, it was decided to explore the phenomenon more thoroughly by presenting subjects with a list of which just over 60% could be pronounced correctly in more than one way.

For a long time, it was assumed that reading theories developed in mainly Anglocentric settings and extended to Latinate languages would apply to all languages (Share, 2008). As research into languages such as Chinese and Japanese as well as Hebrew evolved, these theories began to be challenged. The orthographic depth hypothesis (ODH) has strongly influenced how reading processes may differ in different languages according to how transparent and consistent the correspondences are between graphemes and phonemes. English orthography is considered deeper than Spanish, for example, because of its many irregular sound-spelling correspondences. According to the ODH, Arabic, like Hebrew, has two spellings, a consistent mapping of sound to spelling in its fully vowelised form and a deep orthography used in printed text when short vowels are omitted. This is considered to inhibit prelexical processing (Frost & Bentin, 1992). However, it should be noted that the correspondences between orthography and morphology are consistent, as are the correspondences between morphology and pronunciation, and therefore there is a strong element of consistency which appears to partly contradict the ODH.

Word naming of Hebrew homographs indicated that the ambiguous phonology of unvowelised letter strings increased latencies in word naming (Bentin, Bargai & Katz, 1984). In similar vein, it has been reported that in English, different pronunciations of the same spelling which have different meanings, such as ‘rose’, (‘got up’, ‘flower’), have longer latencies. In the present study, the extent to which ambiguous phonology is problematic has been shown to be less than previously thought. Participants can and do read isolated words. Longer latencies in word naming in the present study were related to the number of possible vowelisations of a letter string.

In Arabic, frequency has been shown to be an influential variable in word naming involving homographs (e.g., Seraye, 2004). Seraye's word naming experiment allowed any legal response to be considered correct, therefore it was not possible to distinguish whether the dominant meaning or a subordinate meaning had been accessed. Such information is important to understanding the influences on word naming latencies. In English, multiple senses which were related were associated with shorter RTs, whereas multiple unrelated meanings resulted in longer RTs (Rodd et al., 2002). The present word naming experiment employed data on the production of name agreement vowelisations of unvowelised letter strings to determine the dominant vowelisation, identified as the correct pronunciation for each unvowelised letter string target for reading, and to examine the impact of variation in name (vowelisation) agreement on response latencies. By identifying the dominant vowelisation as the correct response it was possible to examine the influence of frequency and other psycholinguistic variables on performance in the task because the frequency and other properties of target vowelisations could be used to analyse variation in the latencies of correct responses.

It was predicted that high frequency heterophonic homographs would be named faster than low frequency heterophonic homographs. This was supported, as a unit increase in frequency reduced the response time by 12ms. However, the effects of AoA, together with the *H* statistic, were stronger predictors of latencies. Earlier acquired words were named faster. Words were also named faster as the percentage name agreement across all possible vowelisations for a letter string increased. Frequency effects have been reported in several studies, in Hebrew (e.g., Bentin &

Frost, 1987) and in Arabic (Hermena et al., 2015), although the experiment designs were quite different and neither *H* nor AoA were included in the examined variables.

The predictive ability of *H* reflects both the proportion of subjects agreeing on the dominant vowelisation and meaning, as well as the number of alternative forms for each target word. This appears to be the first time that such a measure has been used with Arabic homographs and, as such, it provides a new perspective on complexity in visual word recognition and reading aloud.

Whilst AoA has been explored in a variety of languages, it does not appear to have been previously included in Arabic studies. Its effect has been assumed to be absent since diglossia does not affect the organisation of morphology in the mental lexicon (e.g., Boudelaa & Marslen-Wilson, 2012). In English (Kuperman et al., 2011) and Spanish (González-Nosti et al., 2014), AoA has been shown to influence response times. Lewis (2006) has cautioned against assuming that results of multiple regression can distinguish the effects of AoA and frequency. Nonetheless, the results found in the present experiment strongly suggest that this is an area worthy of further investigation, especially as it seems to lend some support to the assertion by Frost (2012) that models of reading should be developmental.

8.4 Lexical decision

The lexical decision task used 1,352 of the letter strings in the database (described in Chapter 4) and 1,352 non-word letter strings matched for length with a different 40 native speakers of Arabic. The aim was to test the effects of name agreement, AoA, frequency and length on lexical decision and to establish whether there were any differences between these effects in Arabic, Hebrew, English and other languages. A second aim was to establish whether phonological access was involved in lexical decisions whether or not the word was a heterophonic homograph.

Interestingly, neither of the variables directly related to homography (name agreement and the *H* statistic) significantly predicted decision latencies. Results of the mixed effects regression indicated that AoA, frequency, and then the combined frequency and AoA interaction effects were the strongest predictors of RT. Words with higher frequency are associated with shorter latencies, as is the combined effect of frequency with AoA, which makes a contribution almost as great as frequency alone. As the frequency of the unvowelised letter string increases, the size of the AoA effect increases by -28. Length x frequency is also associated with shorter response times, although it is the least significant of the predictors in this model.

These findings indicate a range of possible explanations. The education system teaches letters, then letters with diacritics to give the three possible pronunciations of each consonant, with hours of repetition and reading aloud single fully vowelised words. Children are then very familiar with a certain vocabulary before progressing to reading several words together, after which they read short sentences and then short paragraphs. Over a period of four to five years, the use of diacritics is reduced until by their mid-teens they are reading unvowelised printed text more quickly than vowelised text. This exposure to Modern Standard Arabic incorporates the experiences that lead to the effects of AoA and frequency.

A connectionist model trained in this way could produce weightings which could explain how words learned early could be stored as whole words with greater weight in the mental lexicon because they formed part of the original network, while words encountered frequently in print could be adding weights within the network. Representations corresponding to words learned later in life and to longer words learned earlier but not met frequently in print would be associated with lower weightings. However, length alone produces an opposite effect; longer words reduce response times. This is probably due to longer words having fewer orthographic or morphological neighbours.

The possible ways in which the mental lexicon is organised could therefore include the orthographic lexicon (letters and diacritics), the phonological lexicon (sounds including the sounds created by a consonant and accompanying diacritic), a morphological lexicon (roots, patterns and possibly CV skeleton), a store of grammatical rules with subsections for different types of verb, noun and inflection, a lexicon of unvowelised letter strings, and a lexicon of whole words. The present study does not provide enough information to propose a design, but highlights the elements which would be needed in a model capable of explaining the different effects.

Ambiguity in English is mainly related to the effects of mappings between orthography and semantics, when there are multiple meanings related to the same spellings sound-spelling inconsistencies (Rodd et al., 2002: Rodd, 2004). Findings from Hebrew may be more relevant to Arabic. These Semitic languages both use abjads and omit short vowels from printed text. When Bentin and Frost (1987) investigated homographs in Hebrew, they proposed that decisions did not necessarily require phonological disambiguation since any difference in latencies was non-significant. This is consistent with the lexical decision results in the present study. Frost (1995) found that differences in latencies for pointed and unpointed words were also non-significant. In most cases, they deduced, the unpointed orthography shared by the different forms of a homograph was sufficient to enable subjects to make a decision. However, other experiments did find evidence of morphological decomposition (see Deutsch et al., 2000).

In Arabic, the results of the present study show that isolated unvowelised homographs are processed as quickly as unambiguous unvowelised words is consistent with the findings of Bentin and Frost (1987). Lexical decision experiments have also shown that the cursive and connected Arabic script contributes to longer latencies (Khateb et al., 2013).

Having discussed the results and compared them with previous research, consideration needs to be given to how they relate to reading and theories of speech production and visual word recognition.

8.5 Results and theories

Investigations of visual word recognition in languages such as Chinese and Hebrew have caused researchers to question whether the Anglocentric models of visual word recognition and reading apply to all languages (Share, 2008). In particular, Frost (2012) has proposed that, based on his research in Hebrew, all models should take into account the specific characteristics of the language. He argues for directing efforts towards a universal theory of reading which incorporates how language is learned, acquired and developed. In contrast, the particular characteristics of Arabic have led Hansen (2014) to propose that a language-specific model of visual word recognition is appropriate. Both scholars acknowledge that initial visual recognition may be common across languages, but that this initial window gives access to different types of information in a variety of ways.

In printed text in Semitic languages, words contain fewer letters, and the root consonants provide clues to the core meaning and the possible word patterns associated with them. The short vowels in the patterns, together with affixes, provide the missing phonological and morphological elements of the words. Arabic orthography is more complex than Hebrew in terms of the number of different shapes and pronunciations that letters can take, depending on their position in a word. In addition, the Arabic script is largely cursive. Moreover, Saiegh-Haddad and Geva (2008) have argued that in order to read Arabic, the need to identify and match consonantal roots with word patterns is not a straightforward morphological operation but involves phonology, syntax and meaning in addition to orthography.

Most of the models of visual word recognition assume that all letters are relatively equal in value in terms of their contribution to the whole word, whether considered individually or as bigrams, and whether or not they can be identified before being arranged into order (Velan & Frost, 2011). In Semitic languages, this is not the case. Boudelaa (2014) has shown through experiments using priming that root consonants act as an ongoing mechanism for identifying possibilities for words, whereas word patterns can only lead to lexical access once they are complete. In Arabic, all letters are not equal. Although Boudelaa (2010) identified that two consonants of the root – the etymon – could prime a word, these two formed a complete morpheme and were not as flexible as open bigrams in determining a word.

In English and other languages which are concatenative rather than agglutinative, letters and morphemes both tend to follow a very similar linear arrangement, so are easily associated. Experiments in Hebrew have shown that priming by roots in the same morphological family as the target word improved recognition (Feldman & Bentin, 1995), while in Arabic, Boudelaa and Marslen-Wilson (2011) found that priming depended on the productivity of the root alone.

In additional to the characteristics of words and previous research, the present study has highlighted the need for a model which can replicate or explain how heterophonic homographs are read and pronounced, and the differing influences of AoA and frequency.

8.5.1 DRC model

The DRC model was originally based on the regular and irregular spellings in English, and assumed a lexical route and a sublexical route to explain how irregular words could be accessed by being stored and recognised as complete units if they had been encountered on enough occasions. The simplest models of interactive activation assume a direct link between separate letters and words by ordering in a linear fashion. In Hebrew and Arabic, letters are ordered by identification of the root consonants, then the full word pattern including short vowels, and any affixes. In order to apply the rules governing the correspondences of letters to sounds, which are needed for word naming, a full form of the word has to be identified. This is necessary because letter pronunciation depends on its position within the word.

Boudelaa (2014) has proposed that in Arabic, roots and word patterns are central to lexical processes involved in understanding speech and printed text, and that decomposition into roots and patterns is therefore obligatory. The morphemes will be linked with functional, morphosyntactic, phonological and semantic information (Boudelaa, 2014, p.48).

The present study has shown no significant effect of name agreement, types or *H* on lexical decision latencies for unvowelised heterophonic homographs and non-homographic words. This is consistent with Boudelaa's Obligatory Morphological Decomposition hypothesis (Boudelaa & Marslen-Wilson, 2011) which states that all words have to be broken down into their morphemes to reach the phonology of a word. However, *H* did have a significant effect on word naming, which implies it takes longer to identify a word ready for pronunciation. Two possibilities are raised. The first is that a fuller search of roots and patterns together with recognition of affixes (in parallel or sequentially) is needed for word naming, while lexical decision requires only recognition of a valid root. The second is that access to phonological information follows access to morphological information. Exceptions to this process could form a second route, as for example in the processing of imported words for example those used in computational linguistics and which are written in Arabic using the letters closest to the sound but lacking root and pattern morphology. One way in which this changes the concept of DRC is illustrated in Figure 22. In practice there would have to be at least three routes if direct access to the lexicon were possible for exception words. However, the present study cannot make this assumption, and so the role of the direct lexical access route in Figure 22 is left open to question.

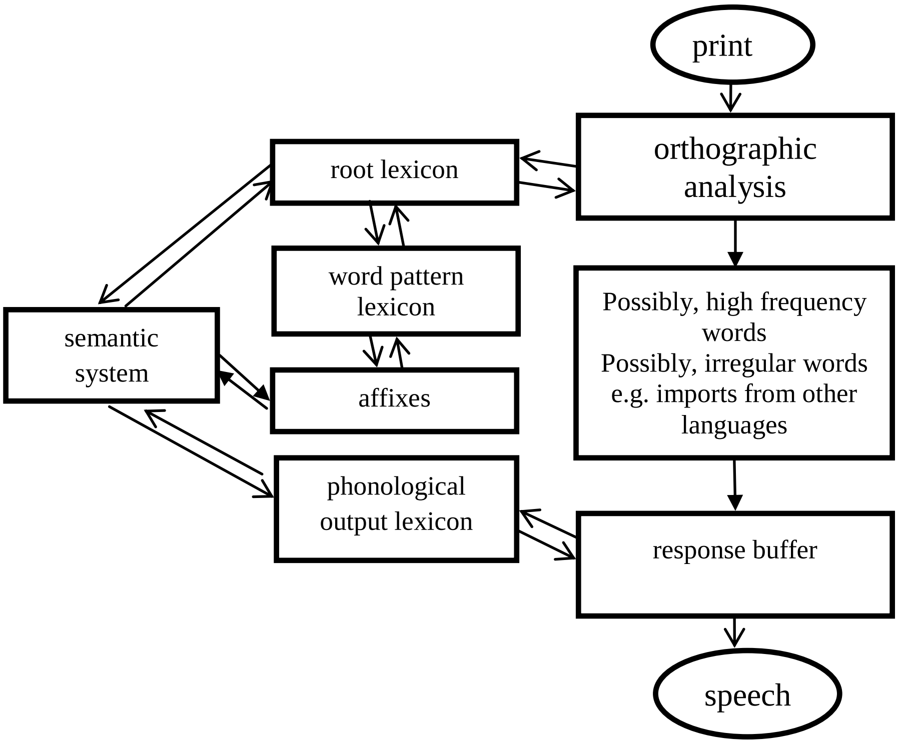


Figure 22 Dual Route revisited

The DRC has been shown to account for frequency (Coltheart, 2005). Similarly, it can account for the effect of *H* found in the present word naming experiment. Response times in the word naming experiment were longer for homographs which had more possible interpretations and greater variability of name agreement. Thus, resolution of the ambiguity took longer. The effect of higher agreement and the effect of frequency can both be explained by the DRC model because high frequency forms of the homographs with high name agreement would be accessed by many individuals more rapidly than low frequency forms.

However, the DRC cannot account so easily for the separate effects of AoA found in the present study. There are various hypotheses regarding the AoA effect. One is arbitrary mapping proposes that less consistency between orthographic, phonological and semantic representations leads to increased AoA and perhaps because early associations will be arbitrary before rules are established (Zevin &

Seidenberg, 2002). Another hypothesis is phonological completeness, which assumes that words learned early are stored as whole words in the lexicon for speech output (Brown & Watson, 1987). A third hypothesis, semantic locus, suggests suggests that words acquired earlier have richer semantic representations which facilitate more rapid access (Brysbaert, Lange, & Van Wijnendaele, 2000).

In a connectionist or neural network, it would not matter which of these was applied because such models assume that rather than having a stored lexicon, the weights of each node on the network are adjusted according to the task (e.g., word naming or lexical decision), the number of times a word is encountered, and additional items of data added as language and vocabulary as acquired. While frequency can be accounted for by models which allow weightings of nodes on a network to change over time as a particular word is encountered repeatedly, AoA is explained by the concept of plasticity (Ellis & Lambon Ralph, 2000). This proposes that words learned at an early age are more strongly entered because the network or mental lexicon was more plastic when it was formed, and that such plasticity is gradually reduced over time.

A connectionist network could also account for the disambiguation of isolated heterophonic homographs and the effects of frequency and AoA. If a heterophonic homograph has many possibilities, it is unlikely that a choice can be made in a sequential manner but instead that there will be several comparisons of the printed letters with the possible word patterns. The dominant meaning will result from a combination of AoA and frequency effects.

Thus, there is no simple answer to which of the main theoretical models best explains the findings of the present study. A connectionist model needs to be able to explain how roots and patterns are matched (for word naming) and requires greater explanation of how the task affects the process, for example to be able to explain why RTs in lexical decision are not affected by homography or why they are affected only by roots and not by patterns. The importance of how Modern Standard Arabic is acquired in childhood indicates that any model of reading aloud, visual recognition or reading which explains the cognitive processes involved in Arabic needs to include developmental aspects. Monaghan and Ellis (2010) have reported how this can be done in English, but the particular implementation of slots in the input layer and hidden layer of phoneme slots would need to be adjusted to take into account the constraints on relative letter-position needed for successful activation of root morpheme units (Velan & Frost, 2011).

Other theories explain some of the findings in the present study but there is no unifying factor. For example, the ODH can explain how, in a deep orthography which does not show short vowels, readers will adopt a visual rather than a spelling to sound correspondence in word naming or lexical decision. However, it cannot explain how readers deal with unfamiliar letter strings, when they have to use their intuitive knowledge of morphology and grammar to reach the appropriate pronunciations. The Obligatory Morphological Decomposition (Boudelaa & Marslen-Wilson, 2011) can explain how a root and word pattern can give access to the required pronunciation of an unvowelised letter string, but cannot explain why the response times in the present lexical decision study showed no effect of homography and the word naming latencies were clearly affected by name agreement.

In many ways, the present study has raised more questions than it has answered. Returning to two of the original broad research questions:

How is reading in Arabic affected by a highly homographic orthography as it appears in printed text?

What are the implications of observations of these effects for understanding the cognitive processes involved?

Firstly, word naming latencies are affected by the number of alternative vowelisations of a specific unvowelised letter string. More possibilities require longer latencies. However, participants in the name agreement study were able to produce correct alternative vowelisations and meanings which provided evidence that they understood by automaticity the morphology and syntax of Modern Standard Arabic. Many of them also showed good vocabulary knowledge in their ability to produce alternative vowelisations. The overall effect of homography in printed Arabic text was less detrimental to reading than the initial literature review suggested. For lexical decision, there was no effect of homography on latencies.

One of the implications of the observations from the four studies is that different cognitive processes are employed when dealing with varying degrees of homography. Some of these processes are different from those seen in the more Anglocentric models, notably the process of accessing non-concatenative morphemes to access pronunciation. A second process relating to spelling-to-sound correspondences is the attachment of short vowels to consonants by the use of diacritics. Short vowels are not learned as sounds which are separate from their consonants. Children spend many hours learning the sounds of consonants with each short vowel. The cognitive process involved in lexical decision in Arabic may consist of identifying a valid root and nothing else, or it may involve identifying the whole unvowelised letter string. It appears there are more processes available for dealing with the particular characteristics of the Arabic language, as it appears in printed MSA text, than in some other languages. This has implications for how the mental lexicon might be structured. All these would benefit from further research using a wider range of psycholinguistic variables, studies and modelling. The present study has opened a door for future studies.

8.5.2 Limitations

Almost all studies have limitations and the present study is no exception. The sample sizes are too small for the results to be generalisable, even though the sample sizes for the word naming and lexical decision tasks are fairly typical of Ph.D. student work. In view of the fact that MSA is the formal language of education, a more detailed breakdown of education level would have been useful, similar to that used by Birchenough, Davies and Connelly (2017).

A more serious limitation is the fact that some potentially important psycholinguistic variables were not controlled for: interference from colloquial Arabic in the word naming study, morphological family size and which letter strings represented homophonic homographs (same spelling, same pronunciation, different meaning). There are many other variables which could be used, but these three relate most directly to Arabic. Controlling for appropriate measures of neighbourhood (orthographical, phonological and morphological) would require very highly skilled Arabic linguists to determine which way to approach this.

There is scope to do much more with the newly created database which has not yet been addressed, in particular analysis of participants' errors in word naming and lexical decision, in order to establish whether errors relate to specific groups of words with particular attributes and whether that could shed further light on the findings of the present study.

8.5.3 Directions for further research

Data from the present study can be further examined in several ways. Firstly, the types of errors made in word naming can be used to investigate whether errors were due to misreading affixes, or selection of an alternative word pattern, or a word with a different frequency. This would yield further insight into the cognitive processes at work. Similarly, analysis of errors in the lexical decision task could shed light on factors leading to the errors, which could be associated with properties of the real words represented by unvowelised letter strings or properties of non-word stimuli, or both. These analyses would strengthen the findings of the present study. r

Equally importantly, the database could be used for a range of further experiments with the aim of better understanding how heterophonic homographs are processed in Arabic. Techniques such as fMRI could be used to investigate whether brain activity differs according to the number of types. Moreover, if the additional data already collected on homophonic homographs were organised by linguistics experts, analysed and used in such experiments, understanding of the processes involved would be even greater. More modestly, the database could be used in experiments such as masked priming, to compare results with previous research using other databases.

Crucially, the differences in cognitive processes which are built into the particular characteristics of the Arabic language merit further investigation.

Bibliography

Abdelhadi, S., Ibrahim, R., & Eviatar, Z. (2011). Perceptual load in the reading of Arabic: Effects of orthographic visual complexity on detection. *Writing Systems Research*, *3*(2), 117-127.

Abdul Razak, Z. R. (2011). *Modern media Arabic: a study of word frequency in world affairs and sports sections in Arabic newspapers*. PhD thesis, University of Birmingham. Retrieved 19 May, 2017, from http://etheses.bham.ac.uk/2882/1/Razak11PhD.pdf

Abu-Hamour, B. (2013). Arabic Reading Comprehension and Curriculum Based Measurement. *International Journal of Humanities and Social Science* 3(17), 184-195.

Abu-Rabia, S. (1995). Learning to read in Arabic: Reading, syntactic, orthographic and working memory skills in normally achieving and poor Arabic readers. *Reading Psychology: An International Quarterly*, *16*, 351–394.

Abu-Rabia, S. (1996). The role of vowels and context in the reading of highly skilled native Arabic readers. *Journal of Psycholinguistic Research, 25*, 629–641.

Abu-Rabia, S. (1997a). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading and Writing: An Interdisciplinary Journal*, *9*, 65–78.

Abu-Rabia, S. (1997b). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers in reading paragraphs, sentences, and isolated words. *Journal of Psycholinguistic Research*, *26*, 465–482.

Abu-Rabia, S. (1998). Reading Arabic texts: Effects of text type, reader type and vowelisation. *Reading and Writing: An Interdisciplinary Journal*, *10*, 105–119.

Abu-Rabia, S. (1999). The effect of Arabic vowels on the reading comprehension of second and sixth-grade native Arab children. *Journal of Psycholinguistic Research*, *28*, 93–101.

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing: An Interdisciplinary Journal*, *14*, 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.

Abu-Rabia, S. (2007). The role of morphology and short vowelisation in reading Arabic of normal and dyslexic readers in Grades 3, 6, 9, and 12. *Journal of Psycholinguistic Research*, *36*, 89-106.

Abu-Rabia, S. (2012). The Role of morphology and short vowelisation in reading morphological complex words in Arabic: Evidence for the domination of the morpheme/root-based theory in reading Arabic. *Creative Education*, *3,* 486-494.

Abu-Rabia, S., & Abu-Rahmoun, N. (2012). The role of phonology and morphology in the development of basic reading skills of dyslexic and normal native Arabic readers. *Creative Education, 3,* 1259-1268. doi: 10.4236/ce.2012.37185.

Abu-Rabia, S., & Awwad, J. S. (2004). Morphological structures in visual word recognition: The case of Arabic. *Journal of Research in Reading*, *27*(3), 321-336.

Abu-Rabia, S., Share, D., & Mansour, M. S. (2003). Word recognition and basic cognitive processes among reading disabled and normal readers in the Arabic language. *Reading and Writing: An Interdisciplinary Journal, 16,* 423-440.

Abu-Rabia, S., & Siegel, L. S. (1995). Different orthographies, different context effects: The effects of Arabic sentence context in skilled and poor readers. *Reading Psychology: An International Quarterly*, *16*, 1–19.

Abu-Rabia, S., & Siegel, L. S. (2003). Reading skills in three orthographies: The case of trilingual Arabic-Hebrew-English-speaking Arab children. *Reading and Writing: An Interdisciplinary Journal, 16*(7), 611-634.

Abu-Rabia, S., & Taha, H. (2004). Reading and spelling error analysis of native Arabic dyslexic readers. *Reading and Writing: An Interdisciplinary Journal, 17*, 651- 689.

Abu-Rabia, S., & Taha, H. (2006). Reading in Arabic orthography: Characteristics, research findings, and assessment. In: R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy*. Mahwah, NJ: Lawrence Erlbaum Associates, pp.321-338

Adelman, J. S. (2011). Letters in time and retinotopic space. *Psychological Review, 118*, 570-582

Adelman, J. S. (ed.) (2012). *Visual word recognition volume 2: Meaning and context, individuals and development.* Hove: Psychology Press

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive Psychology, 68,* 113–160.

Ahmad, H. A., Ibrahim, R., & Share, D. L. (2014) cognitive predictors of early reading ability in Arabic: A longitudinal study from kindergarten to Grade 2 In E. Saiegh-Haddad, & M. Joshi (Eds.). *Handbook of Arabic Literacy: Insights and perspectives* (pp.171-196). Dordrecht: Springer.

Alario, F-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods Instruments & Computers,* 36(1), 140-155.

Al Fawzaan, A., Hussein, M., & Fadh, M. (2003) *Al`Arabiyyatu Baina Yadaik* Riyadh: Islamic Charity Association

Al Jarf, R. (2007). *A contrastive analysis of English and Arabic morphology for translation students.* Course material. https://www.academia.edu/14942469/A_Contrastive_Analysis_of_English_and_Arabic_Morphology_for_Translation_Students

Al-Dahdah, A. (1989). *The grammar of the Arabic language in tables and lists.* Beirut: Maktabat Lebanon (in Arabic)

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, & Cognition,***15**, 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 234-254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*(4), 439-461.

Andrews, S., & Davis, C. (1999). Interactive Activation Accounts of Morphological Decomposition: Finding the Trap in Mousetrap? *Brain and Language, 68,* 355–361.

Armstrong, B. C., & Plaut, D.C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience, 31*, 940-996

Attia, M. A. (2008). *Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation.* Unpublished PhD thesis, University of Essex. Retrieved 16 May, 2017, from orb.essex.ac.uk/lg/lg617/restricted/Attia-PhD-Thesis.pdf

Ayari, S. (1996). Diglossia and illiteracy in the Arab world. *Language, Culture and Curriculum*, 9, 243–252.

Azzam, R. (1993). The nature of Arabic reading and spelling errors of young children. *Reading and Writing*, 5, 355–385.

Baayen, H. (1994). Derivational productivity and text typology. *Journal of Quantitative Linguistic, 1*, 16-34.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance. 10*, 340-57

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords.* St. Louis: Washington University. Available at elexicon.wustl.edu.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S.D., Spieler, D.H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

Balota, D. A., Law, M. B., & Zevin, J. D. (2000). The attentional control of lexical processing pathways: Reversing the word frequency effect. *Memory & Cognition, 28*(7), 1081-1089.

Balota, D. A., Paul, S. T., & Spieler, D. H. (1999). Attentional control of lexical processing pathways during word recognition and reading. In S. Garrod & M. Pickering (Eds.). *Language processing.* pp. 15-57. London: Psychology Press.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639-647.

Balota, D. A, & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: beyond measures of central tendency. *Journal of Experimental Psychology: General*. 128: 32-55.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update) In Traxler, M. & Gernsbacher, M.A. (eds.) (2006) *Handbook of Psycholinguistics* (pp.285-376). Oxford: Academic Press.

Balota, D. A., Yap, M. J., Cortese, M.J., Hutchison, K. A., Kessler, B., Loftis, B… et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445-45

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 9*, 495–523.

Baluch, B. (1990). Word recognition in Persian. *Baranoosh Scientific Quarterly*, *1*, 3–6.

Baluch, B. (1993). Lexical decisions in Persian: A test of the orthographic depth hypothesis. *International Journal of Psychology, 28*, 19–29.

Baluch, B. (2005). *Persian orthography and its relation to literacy.* In R. M. Joshi, & P.G. Aaron (Eds.), *Handbook of Orthography and literacy* (pp. 365-376). London: Lawrence Erlbaum Associates.

Baluch, B., & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 644–652.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*(3), 373–400.

Barry, C., Morrison, C.M., & Ellis, A.W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology, 50A*, 560-585.

Bauer, D.J., & Curran, P.J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*, 373-400.

Benmamoun, E. (2003). The role of the imperfective template in Arabic morphology. In J. Shimron (Ed.), *Language processing and acquisition in languages of Semitic root-based, morphology* (pp.99–114). Amsterdam: John Benjamins.

Bentin, S., & Frost, R, (1987). Processing lexical ambiguity and visual word recognition in a deep orthography. *Memory & Cognition, 15*(1), 13-23.

Bentin, S., & Ibrahim, R. (1996). New evidence for phonological processing during visual word recognition: The case of Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 309-323.

Berent, I. (1997). Phonological priming in the lexical decision task: Regularity effects are not necessary evidence for assembly. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1727–1742.

Besner, D., & Humphreys, G. (Eds.). (2009). *Basic processes in reading: Visual word recognition*. New York, NY: Routledge

Binti Faizal, S., Khattab, G., & McKean, C. (2015). *The Qur'an Lexicon Project: A database of lexical statistics and phonotactic probabilities for 19,286 contextually and phonetically transcribed types in Qur'anic Arabic*. Retrieved http://eprint.ncl.ac.uk/file_store/production/213017/D0E1A4EA-A847-41BC-98D3-C7149E1ECFD2.pdf

Birchenough, J M. H., Davies, R., & Connelly, V. (2017). Rated age-of-acquisition norms for over 3,200 German words. *Behavior Research, 49*,484–501

Black, S. R. (2001). Semantic satiation and lexical ambiguity resolution. *American Journal of Psychology, 114*(4), 493-510.

Blythe, H. I. (2014). Developmental changes in eye movements and visual information encoding associated with learning to read. *Current Directions in Psychological Science, 23*(3), 201–207.

Blythe, H. I., Pagán, A., & Dodd, M. (2014). Beyond decoding: phonological processing during silent reading in beginning readers. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition, 40*(6) http://dx.doi.org/10.1037/xlm0000080. Retrieved 15 May, 2017, from https://www.researchgate.net/profile/Hazel_Blythe/publication/268213432_Beyond_Decoding_Phonological_Processing_During_Silent_Reading_in_Beginning_Readers/links/555afca908ae6943a877e9c7.pdf.

Bohas, G. (2006). The organization of the lexicon in Arabic and other Semitic languages. In Boudelaa, S. (Ed.) *Perspectives on Arabic Linguistics XVI* (pp.1-37). Amsterdam: John Benjamins Publishing Company.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*, 89-114.

Boudelaa, S. (2014). Is the Arabic mental lexicon morpheme-based or stem-based? Implications for spoken and written word recognition. In E. Saiegh-Haddad, & R.M. Joshi (Eds.). *Handbook of Arabic Literacy. Insights and Perspectives.* Series: Literacy Studies, Vol. 9. (pp.31-54). London: Springer.

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition, 81*, 65-92.

Boudelaa, S., & Marslen-Wilson, W. D. (2004a). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, *90*, 106-116.

Boudelaa, S., & Marslen-Wilson, W. D. (2004b). Abstract morphemes and lexical representation: The CV-skeleton in Arabic. *Cognition, 92,* 271-303.

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, *20*, 207-260.

Boudelaa, S. & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behavior Research Methods*, *42*(2), 481-487.

Boudelaa, S. & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes, 26*(4), 624-652.

Boudelaa, S. & Marslen-Wilson, W. D. (2013). Morphological structure in the Arabic mental lexicon: Parallels between standard and dialectal Arabic. *Language and Cognitive Processes*, *28*(10), 1453–1473.

Boudelaa, S., & Marslen-Wilson, W. D. (2015). Structure, form, and meaning in the mental lexicon: evidence from Arabic. *Language, Cognition and Neuroscience*, *30*(8), 955–992. http://doi.org/10.1080/23273798.2015.1048258

Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. D. (2009). Arabic morphology in the neural language system: A mismatch negativity study. *Journal of Cognitive Neuroscience*, *22*, 998–1010. 10.1162/jocn.2009.21273

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), 585-599. doi: 10.3758/s13428-015-0602-3

Bourisly, A. K., Haynes, C., Bourisly, N., & Mody, M. (2013). Neural correlates of diacritics in Arabic: An fMRI study. *Journal of Neurolinguistics, 26*, 195–206.

Breznitz, Z. (2004). Introduction on regular and impaired reading in semantic languages. *Reading and Writing: an Interdisciplinary Journal*, *17*, 645-649.

Brown, G. D. & Watson, F. L. (1987) First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory &: Cognition*, *15*(3), 208-216.

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology, 64,* 545–559

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*, 992–1011.

Brysbaert, M., & New, B. (2009) Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Brysbaert, M., Lange, M., & Van Wijnendaele, M. L. I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, *12*(1), 65-85.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412-424.

Buckwalter, T. (2002). *Buckwalter Arabic Morphological Analyzer, Version 1.0.* Philadelphia: Linguistic Data Consortium, Catalog No. LDC2002L49, ISBN 1-58563-257-0.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297-332.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, *12,* 590-606.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.). *Strategies of information processing* (pp.151- 216). New York: Academic Press.

Coltheart, M. (2005). Modeling reading: The dual-route approach. In: M.J. Snowling & C. Hulme (eds). *The Science of Reading,* (pp.6-23). Malden: Blackwells Publishing. doi: 10.1002/9780470757642.ch1

Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *London Review of Education*, *4,* 5-17. DOI 10.1080/13603110600574322

Coltheart, M. (2012). Dual-route theories of reading aloud. In J. S. Adelman (Ed.), *Visual Word Recognition Volume 1: Models and Methods, Orthography and Phonology* (illustrated ed.), (pp. 3-27). UK: Psychology Press.

Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & Performance VI,* (pp. 535-555). Hillsdale, NJ: Erlbaum.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1197-1211

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204-256.

Cortese, M. J., & Balota, D. A. (2012). Visual word recognition in skilled adult Readers. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 159-185). New York: Cambridge University Press.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60*, 1072–1082.

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*, *40*, 791– 794.

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition, 28*, 1269-1276.

Cramer, P. (1970). A study of homographs. In Postman, L. & Keppel, G. (Eds.), *Norms of word association*, (pp. 361-370). New York: Academic Press.

Cuetos, F., Ellis, A.W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers, 31*,650-658.

Cutler, A., & Davis, C. (2012). An orthographic effect in phoneme processing, and its limitations. Frontiers in Psychology, 3, 18. doi:10.3389/fp

Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & Cognition, 41,* 297–311

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*, 713-758.

Davoudi, M., & Moghadam, H.R.H. (2015). Critical review of the models of reading comprehension with a focus on situation models. *International Journal of Linguistics, 7*(5), doi:10.5296/ijl.v7i5.8357.

Deacon, S. H., & Kirby, J. D. (2004). Morphological awareness: Just “more phonological”? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics, 25*, 223–238. DOI: 10.1017.S0124716404001117. Retrieved 15 May, 2017, from https://www.researchgate.net/profile/John_Kirby5/publication/231747380_Morphological_awareness_Just_more_phonological_The_roles_of_morphological_and_phonological_awareness_in_reading_development/links/00b7d5182828b5e0e8000000.pdf

De Groot, A. (2013). Reading. In Grosjean , F. & Li, P. (Eds.),*The Psycholinguistics of Bilingualism* (p.73-99). Oxford: Wiley-Blackwell.

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: lexical access and grammatical encoding, *Cognitive Science, 23*(4), 517-542.

Dell’Acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers, 32*, 588-615

Deutsch, A. & Frost, R. (2003). Lexical organization and lexical access in a non-concatenated morphology: Mapping the mental lexicon. In Y. Shimron (Ed.), *Language processing and acquisition in Languages of Semitic, Root Based Morphology.* Amsterdam: John Benjamins

Deutsch, A., Frost, R. & Forster, K. I. (1998). Verbs and nouns are organized and accessed differently in the mental lexicon: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1238-1255.

Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2000). Early morphological effects in word recognition in Hebrew: Evidence from parafoveal preview benefit. *Language and Cognitive Processes, 15***,** 487-506.

Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Language and Cognitive Processes, 14,* 487-506.

Ellis, N.C., & Hooper, A.M. (2001). It is easier to learn to read in Welsh than in English: Effects of orthographic transparency demonstrated using frequency-matched cross-linguistic reading tests. *Applied Psycholinguistics*, *22*, 571–599.

Ellis, A.W., & Lambon Ralph, M.A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1103-1123.

Ellis, A.W., & Morrison, C.M. (1998). Real age of acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 515-523.

Eysenck, M. W. & Keane, M. T. (2015). Reading and speech perception. In *Cognitive Psychology: A Student's Handbook,* (7th Ed.). (pp. 353-402). Hove: Psychology Press

Farghaly, A. & Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *Transactions on Asian Language Information Processing (TALIP), the Association for Computing Machinery (ACM), 8*(4), 1-22.

Feldman, L. B., Pnini, T., & Frost, R. (1995). Decomposing words into their constituent morphemes. Evidence from English and Hebrew. *Journal of Experimental Psychology, 21*, 947–960.

Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *47A*(3), 365-382.

Ferrand, L., & New, B. (2003). Mental lexicon: Some words to talk about words. In P. Bonin (Ed.), *Mental lexicon: Some words to talk about words*. Hauppage, NY: Nova Science Publishers.

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology*, 2:306. doi:10.3389/fpsyg.2011.00306

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124.

Frisch, S., & Zawaydeh, B. (2001). The psychological reality of OCP-Place in Arabic. *Language*, *77*(1), 91-106.

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *20*, 1-16.

Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. J*ournal of Experimental Psychology: Learning, Memory and Cognition, 21*, 398-408.

Frost, R. (2006). Becoming literate in Hebrew: the grain size hypothesis and Semitic orthographic systems *Developmental Science, 9* (5), 439-440.

Frost, R. (2009). Reading in Hebrew vs. reading in English: Is there a qualitative difference? In K. Pugh & P. McCardle (Eds.) *How Children Learn To Read: Current Issues and New Directions in the Integration of Cognition, Neurobiology and Genetics of Reading and Dyslexia Research and Practice*, (pp. 235-254). Hove: Psychology Press.

Frost, R. (2012). Towards a universal model of reading *Behavioral Brain Science, 35*(5), 263–279.

Frost, R. & Bentin, S. (1992). Processing phonological and semantic ambiguity: Evidence from semantic priming at different SOAs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 58-68.

Frost, R., Deutsch, A., Gilboa, O., Tannebaum, M., & Marslen-Wilson, W. D. (2000). Morphological priming: Dissociation of phonological, semantic and morphological factors. *Memory and Cognition*, *28*, 1277–1288.

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 829–856.

Frost, R., Grainger, J. &. Rastle, K. (Eds.). (2005). Current Issues in Morphological Processing. In: *Special Issue of Language and Cognitive Processes*. Hove, Psychology Press.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-115.

Ganayim, D. (2015). Letter visibility and the optimal viewing position effect of isolated connected and un-connected letters in Arabic. *Psychology of Language and Communication,* 19(3). DOI: https://doi.org/10.1515/plc-2015-0011

Gaskell, M. G. & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45,* 220-266.

Gawlick-Grendell, L. A., & Woltz, D. J. (1994) Meaning dominance norms for 120 homographs. *Behavior Research Methods, Instruments, & Computers*, 26(1), 5-25.

Gee, N. R., & Harris, S. L. (2010). Homograph norms: An alternative approach to determining meaning dominance. *Behavior Research Methods, 42*(4), 976-986.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*, 395– 427.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human perception and performance*, *5*(4), 674-691.

González-Nosti, M., Barbón, A., Rodríguez-Ferreiro. J., & Cuetos, F. (2014). Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2,765 words. *Behavior Research Methods, 46*(2), 517-525.

Gorfein, D. S., Berger, S., & Bubka, A. (2000) The selection of homograph meaning: Word association when context changes *Memory & Cognition 28*(5), 766-773.

Gorfein, D. S., Viviani, J. M., & Leddo, J. (1982) Norms as a tool for the study of homography. *Memory & Cognition 10*(5), 503-509.

Grainger, J., & Ziegler, J. C. (2011). A Dual-Route Approach to Orthographic Processing. *Frontiers in Psychology*, *2*, 54. Retrieved from http://doi.org/10.3389/fpsyg.2011.00054

Grainger, J., & van Heuven, W.J.B. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.) *Mental lexicon: "Some words to talk about words"*. New York: Nova Science Publishers

Grindrod, C., & Baum, S. (2003). Sensitivity to local sentence context information in lexical ambiguity resolution: Evidence from left-right-hemisphere-damaged individuals, *Brain & Language*, *85*, 502-523.

Grondelaers, S., & Speelman, D. (2007). A variationist account of constituent ordering in presentative sentences in Belgian Dutch. *Corpus Linguistics and Linguistic Theory*, *3*(2), 161-193

Habash, N.Y. (2010) *Introduction to Arabic Natural Language Processing* in series Ed Hirst, G. (Synthesis Lectures on Human Language Technologies), San Rafael, CA: Morgan and Claypool Publishers

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review*, *111*, 662-720.

Hautala, J., Hyönäc, J. & Arob, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns *Vision Research 51*(15), 1719–1727

Havelka, J., & Tomita, I. (2006). Can distributed orthographic knowledge support word-specific long-term priming? *Visual Cognition*, *13*(7-8), 981-991

Haywood, J. A., & Nahmad, H. M. (1962) *A new Arabic grammar of the written language.* Lund Humphries

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

Hino, Y., Lupker, S. J., & Pexman, P. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2002, 28*(4), 686–713

Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience and Biobehavioral Reviews, 46*, 85–104

Holes, C. (2004). *Modern Arabic. Structures, Functions, and Varieties*. Revised Edition. Georgetown Classics in Arabic Language and Linguistics. Washington, D.C.: Georgetown University Press

Hu, M., & Nation, I. S. P. (2000). Vocabulary Density and Reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430

Ibrahim, R. & Aharon-Peretz, J. (2005). Is literary Arabic a second language for native Arab speakers?: Evidence from a semantic priming study. *The Journal of Psycholinguistic Research*, 34(1),51-70

Ibrahim, R., Eviatar. Z., & Aharon-Peretz, J. (2007). Metalinguistic awareness and reading performance: A cross language comparison. *The Journal of Psycholinguistic Research*, *36*(4), 297-317

Ibrahim, R. (2008). Does visual and auditory word identification have a language-selective input? *The Journal of Psycholinguistic Research*, *36*(4), 297-317.

Ibrahim, R. (2009a). Selective deficit of second language: a case study of a brain-damaged Arabic-Hebrew bilingual patient. *Behavioral and Brain Functions : BBF*, *5*, 17. http://doi.org/10.1186/1744-9081-5-17.

Ibrahim, R. (2009b). The cognitive basis of diglossia in Arabic: Evidence from a repetition priming study within and between languages *Psychology Research and Behavior Management, 12*, 95-105

Ibrahim, R. (2010). Vowels in Semitic alphabet languages. In E. F. Caldwell (Ed.), *Bilinguals: Cognition, Education and Language Processing* (pp.147-166), New York, Nova Science Publishers Inc

Ibrahim, R. (2013). Reading in Arabic: New Evidence for the Role of Vowel Signs. *Creative Education*, *4*, 248-253

Ibrahim, R., Eviatar, Z., & Aharon-Peretz, J. (2002). The characteristics of Arabic orthography slows its processing. *Neuropsychology*, *16*, 322-326

Idrissi, A., & Kehayia, E. (2004). Morphological units in the Arabic mental lexicon: Evidence from an individual with deep dyslexia. *Brain and Language, 90*(1-3), 183–197

Iraqi, J. (1990). *Reading to Arabic-speaking kindergarten children compared to alternative enrichment activities as a means of improving listening comprehension and language skills*. Unpublished Master's Thesis, University of Haifa, Haifa, Israel

Izre'el, S. (2003). The Emergence of Spoken Israeli Hebrew. In: Benjamin H. Hary (ed.). *Corpus Linguistics and Modern Hebrew: Towards the Compilation of The Corpus of Spoken Israeli Hebrew* (CoSIH). Tel Aviv, Tel Aviv University, The Chaim Rosenberg School of Jewish Studies. pp.85-104

Jones, L. L. & Estes, Z. (2012). Lexical Priming: Associative, Semantic, And Thematic Influences On Word Recognition in Adelman, J. (ed.) (2012) *Visual Word Recognition Volume 2: Meaning and Context, Individuals and Development* Hove, Psychology Press p.44-72.

Jordan, T. R., Almabruk, A. A. A., McGowan, V. A., & Paterson, K. B. (2011). Evaluating hemispheric divisions in processing fixated words: The evidence from Arabic. *Cortex, 47*, 992-997.

Jordan, T. R., Paterson, K. B., & Almabruk, A. A. A. (2010). Revealing the superior perceptibility of words in Arabic. *Perception*, 39, 426-428

Jordan, T. R., Paterson, K. B., Kurtev, S. & Xu, M. (2010). Re-evaluating split-fovea processing in word recognition: effects of word length during monocular viewing. *Cortex, 46*(1), 100-105.

Joshi, R. M. (2005). Vocabulary: A critical component of comprehension. *Reading and Writing Quarterly, 21*, 209-219.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131*, 684–712.

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. H. (2012). What lexical decision and naming tell us about reading. *Reading and Writing, 25*(6), 1259–1282

Katz. L., & Frost, R, (1992). Reading in different orthographies: the orthographic depth hypothesis. In R. Frost, & L. Katz (Eds.), *Orthography, Phonology Morphology, and Meaning* pp. 67-84.

Koriat, A. (1984). Reading without vowels: Lexical access in Hebrew. In H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and Performance X: Control of language processes* (pp. 227-242). Mahwah, New Jersey: Lawrence Erlbaum Associates

Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. (2012) Age-of-acquisition ratings for 30,000 English words. *Behavior Resarch Methods, 44*(4), 978-990.

Khwaileh, T., Body, R., & Herbert, R.J. (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. doi: 10.1007/s10936-013-9277-z.

Lavidor, M., & Whitney, C. (2005). Word length effects in Hebrew. *Cognitive Brain Research, 24*, 127-32.

Lebanese Arabic Institute (2017). *Arabic Alphabet Chart*. Retrieved from https://www.lebanesearabicinstitute.com/arabic-alphabet.

Lewis, M. P., Simons, G. F. & Fennig, C. D. (2014). *Ethnologue: Languages of the World*, 17th Ed., Dallas, Texas, SIL International.

Library of Congress (n.d.) *ALA-LC Romanization Tables* Retrieved from https://www.loc.gov/catdir/cpso/roman.html

Liu, Y., Shu, H. & Li, P. (2007) Word naming and psycholinguistic norms: Chinese *Behavior Research Methods, 39*(2), 192-198.

Luce, P. A., & Pisoni, D. B. (1998) Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1–36.

Lupker, S. J. (2005) Visual Word Recognition: Theories and Findings. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp.39-60). Oxford: Blackwell Publishing Ltd.

Mahfoudhi, A. (2007). The place of the etymon and the phonetic matrix in the Arabic mental lexicon. *Journal of Arabic and Islamic Studies 7*, 74-102.

Marian, V., & Blumenfeld, H.K. (2006) Phonological Neighborhood Density Guides: Lexical Access in Native and Non-Native Language Production *Journal of Social and Ecological Boundaries 2*(1), 3-35.

Marslen-Wilson, W.D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Process 23*(3), 394-421.

Martin, C., Vu., H., Kellas, G., & Metcalf, K. (1999). Strength of discourse contexts as a determinant of the subordinate bias effect. *Quarterly Journal of Experimental Psychology, 52A*(4), 813-839.

McBride-Chang, C. (2004) *Children's literacy development*. New York, Oxford University Press

McCann, R.S., & Besner, D. (1987) Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 14

McClelland, J., & Rumelhart. D. (1981) An interactive activation model of context effects in letter perception: part 1. An account of basic findings. *Psychol. Rev*. *88*,375–407

McClelland, J.L., Rumelhart, D.E., & the PDP research group (1986), *Parallel Distributed Processing*, Volume II, Cambridge, Mass., MIT Press

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *6*, 1139-1189. DOI: 10.1111/cogs.12146

McDonald, S.A (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length *Visual Cognition 13*(1), 89-98

Miller, P., Liran-Hazan, B., & Vaknin V.J. (2016). Morphological Decomposition in Reading Hebrew Homographs *Journal of Psycholinguistics Research 45*(3), 717-738.

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63*, 506–525.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology, 50A*, 528-559.

Morrison, C.M., & Ellis, A.W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology, 91*, 167-180.

Myhill, J. (2014). The Effect of Diglossia on Literacy in Arabic and Other Languages. E. Saiegh-Haddad, R. M. Joshi (Eds.), *Handbook of Arabic Literacy* (pp. 197-223). Dordrecht: Springer.

Nelson, D.L., McEvoy, C.L., & Schreiber, T.A (2004). The University of South Florida free association, rhyme, and word fragment norms *Behavior Research Methods, Instruments, & Computers 36*(3), 402-407.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(4), 661-677

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review,* 13(1), 45-52

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327-357.

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Science, 17*(10), 517-24.

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General, 137*(3), 434-455.

Norris D., & Kinoshita S. (2012). Reading through a noisy channel: why there's nothing special about the perception of orthography. *Psychological Review, 119*, 517–545.

Noujoumian, P. (2011) *Towards the development of an automatic diacritizer for the Persian orthography based on the Xerox finite state transducer*. PhD thesis, University of Ottowa. Available online at https://www.ruor.uottawa.ca/bitstream/10393/20158/1/Nojoumian_Peyman_2011_thesis.pdf

Onifer, W., & Swinney. D.A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias *Memory & Cognition 9*(3), 225-236.

Owens, J. (2013). A house of sound structure, of marvelous form and proportion: An introduction. In J. Owens (Ed.)., *The Oxford Handbook of Arabic Linguistics.* (pp.1-22). Oxford: Oxford University Press

Paterson, K. B., Almabruk, A.A.A., McGowan, V.A., White, S. J., & Jordan, T. R. (2015) Effects of word length on eye movement control: The evidence from Arabic *Psychonomic Bulletin & Review, 21,* 505-511

Paulston, C.B., Kiesling, S.F., & Rangel, E.S. (2012). *Handbook of Intercultural Discourse and Communication* Chichester, Wiley-Blackwell

Peleg, O., Eviatar, Z., Hazan, H., & Manevitz, L. (2007) Differences and interactions between cerebral hemispheres when processing ambiguous homographs. In Carbonell, J.G. and Siekmann, J. (Eds.) *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint* Volume 4840 of the series Lecture Notes in Computer Science, Berlin, Springer, pp 367-380

Peleg, O. & Eviatar, Z. (2008). Hemispheric sensitivities to lexical and contextual information: Evidence from lexical ambiguity resolution. *Brain and Language*, 105(2), 71-82.

Peleg, O., & Eviatar, Z. (2009). Semantic asymmetries are modulated by phonological asymmetries: Evidence from the disambiguation of homophonic versus heterophonic homographs. *Brain and Cognition*, *70*(1), 154-162.

Peleg, O., Markus, A., & Eviatar, Z. (2012). Hemispheric asymmetries in meaning selection: Evidence from the disambiguation of homophonic vs. heterophonic homographs *Brain and Cognition 80*, 328–337.

Perea, M. (2015). Neighborhood effects in visual-word recognition and reading. In A. Pollatsek and R. Treiman (Eds.) *The Oxford Handbook on Reading* (pp. 76-87). New York, Oxford University Press. DOI: 10.1093/oxfordhb/9780199324576.013.7

Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics*, *67,* 301–314.

Perry, J. R. (2007) Morphology of Persian In Kaye, A. (Ed.) *Morphologies of Asia and Africa*, Vol. 2, (pp. 975-1019) Lake, Ind.: Eisenbrauns,

Pexman P. (2012). Meaning-based influences on visual word recognition. In Adelman J. S. (Ed.), *Visual word recognition* volume 2 (pp. 24–43). Hove, England: Psychology Press

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56

Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of “neighborhood size” in reading and lexical decision. *Journal of Experimental Psychology: Human Perception & Performance, 25*, 1142-1158.

Pollatsek, A. & Treiman, R. (2015). *The Oxford Handbook of Reading* Oxford, Oxford University Press

Pritchard, S. (2013). *A computational model of learning to read via self-teaching*. Unpublished PhD., University of Macquarie, Australia. https://www.researchgate.net/publication/268391458

Protopapas, A. (2007) CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods 39*((4), 859-862.

Rahbari, N. & Sénéchal, M. (2009) Lexical and nonlexical processes in the skilled reading and spelling of Persian *Reading and Writing 22*(5), 511-530.

Rastle, K., Davis, M.H., Marslen-Wilson, W.D., & Tyler, L.K. (2000) Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(4/5), 507–537.

Rastle, K., Davis, M.H., & New, B. (2004). The broth in my brother’s brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review, 11*, 1090–1098.

Ratcliffe, R. (2013). Arabic morphology. In J. Owens (Eds.), *The Oxford handbook of Arabic linguistics*, (pp. 72–91). Oxford: Oxford University Press.

Ravid, D.D. (2012) *Spelling Morphology: The Psycholinguistics of Hebrew Spelling* New York, Springer

Rayner, K., & Reichle, E. D. (2010). Models of the Reading Process. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(6), 787–799. http://doi.org/10.1002/wcs.68

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275-280.

Rodd, J. M. (2004a). The effect of semantic ambiguity on reading aloud: A twist in the tale. *Psychonomic Bulletin & Review, 11*(3), 440-445 doi: 10.3758/BF03196592.

Rodd, J. M. (2004b). When do leotards get their spots? Semantic activation of lexical neighbors in visual word recognition. *Psychonomic Bulletin & Review, 11*(3), 434-439. doi:10.3758/BF03196591.

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language, 87,* 16-37.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2000). The advantages and disadvantages of semantic ambiguity. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*. (pp. 405-410). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: semantic competition in lexical access. *Journal of Memory and Language, 46*, 245—266.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*, 89–104.

Rodd, J. M., Lopez Cutrin, B., Kirsch, H., Millar, A., & Davis, M.H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68*, 180-198. DOI: 10.1016/j.jml.2012.08.002

Rumelhart, D. E. & McClelland, J. L. (1987) Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. Mcclelland, D. E. Rumelhart, & the PDP Research Group (Eds.) *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (p.216-271). Cambridge, MA: MIT Press

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review 89* (1), 60-94

Rumelhart, D.E., Hinton, G E., & Williams, R.J. (1986) Learning internal representations by error propagation. In D.E. Rumelhart & J.L. McClelland (Eds.). *Parallel. Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1,* (pp.318-362). Cambridge, Mass, USA: MIT Press, Ryding, K.C. (2005). *A Reference Grammar of Modern Standard Arabic* (pp.6-7), Cambridge: Cambridge, University Press

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics*, *24*, 431-451.

Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudo-words in a diglossic context. *Applied Psycholoinguistics*. *25*, 495–512

Saiegh-Haddad, E. (2005). Correlates of reading fluency in Arabic: Diglossic and orthographic factors. *Reading and Writing: An Interdisciplinary Journal*, 18, 559-582

Saiegh-Haddad, E. (2013). A tale of one letter: Morphological processing in early Arabic spelling. *Writing Systems Research*, 5(2), 1-20. http://dx.doi.org/10.1080/17586801.2013.857586

Saiegh-Haddad, E., & Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English–Arabic bilingual children. *Reading and Writing*, *21*(5), 481-504.

Saiegh-Haddad, E. & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad, & M. Joshi (Eds.) *Handbook of Arabic Literacy: Insights and perspectives* (pp.3-30) Dordrecht: Springer,

Saiegh-Haddad, E., Levin, I., Hende, N., & Ziv, M. (2011). The Linguistic Affiliation Constraint and phoneme recognition in diglossic Arabic. *Journal of Child Language, 38*(2), 297-315.

Salah, S. (2008). *The relationship between vocabulary knowledge and reading comprehension of authentic Arabic texts* (Master of Arts abstract, Brigham Young University, 2008).

Salim, J.A. (2013) Homonymy in Jordanian Colloquial Arabic: A Semantic Investigation *English Language and Literature Studies* Vol.3, no.3, p.69-76

Schiff, R. (2012). Shallow and deep orthographies in Hebrew: the role of vowelisation in reading development for unvowelised scripts. *Journal of Psycholinguist Research, 41*(6), 409-24.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L.B. Feldman (Ed.), *Morphological aspects of language processing* (p.131–154). Hillsdale, NJ: Erlbaum.

Seidenberg, M.S. (2005). Connectionist Models of Word Reading. *Current Directions in Psychological Science 14*(5), 238-242

Seidenberg, M.S. (2007). Connectionist Models of Reading In Gareth Gaskell. (Ed.), *The Oxford Handbook of Psycholinguistics* p.235-250. Oxford, Oxford University Press

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523.

Seidenberg, M. S., Waters, G. S., Barnes, M., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning & Verbal Behavior, 23,* 383-404.

Seraye, A. M. (2004). *The role of short vowels and context in the reading of Arabic, comprehension and word recognition of highly skilled readers*. Unpublished doctoral dissertation, University of Pittsburgh. Retrieved from http://d-scholarship.pitt.edu/9821/1/AbdullahSeraye.pdf

Seroussi, B. (2011). *The morphology-semantics interface in the mental lexicon: the case of Hebrew.* Unpublished PhD thesis, University of Tel Aviv. Retrieved 25 May, 2017, from humanities1.tau.ac.il/linguistics_eng/images/stories/Batia_Seroussi_PhD_2011.pdf

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*(2). 159-187.

Share, D. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an ‘‘outlier’’ orthography. *Psychological Bulletin, 134*(4), 584–615

Share, D., & Bar-On, A. (2017). Learning to read a semitic abjad: The triplex model of Hebrew reading development. *Journal of Learning Disabilities* Ahead of Print. DOI: 10.1177/0022219417718198

Share, D.L., & Levin, I. (1999). Learning to read and write in Hebrew. In M Harris, and G. Hatano (Eds.), *Learning to read and write,* pp.89-111, Cambridge: Cambridge University Press

Shimron, J., & Sivan, T. (1994). Reading proficiency and orthography: Evidence from Hebrew and English. *Language Learning, 44,* 5-27

Simpson, G.B. and Burgess, C. (1985) Activation and Selection Processes in the Recognition of Ambiguous Words. *Journal or Experimental Psychology: Human Perception and Performance, 11*(1), 28-39.

Simpson, G.B. & Krueger, M.A. (1991) Selective access of homograph meanings in sentence context *Journal of Memory and Language, 30* (6), 627–643

Singson, M., Mohany, D., & Mann, V. (2000). The relation between reading ability and, morphological skills: Evidence from derivational suffixes. *Reading and Writing*, *12*, 219–252.

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. doi:10.1037/0278-7393.6.2.174

Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging*, *15*, 225-231

Taha, H. (2013). Investigating cognitive processes underlying reading in Arabic: Evidence from typical and poor reading performance. *Psychology, 4*, No.12, 1018-1026

Taha, H., & Saiegh-Haddad, E. (2017). Morphology and spelling in Arabic: Development and interface. *Journal of Psycholinguist Research*, *46*, 27–38 DOI 10.1007/s10936-016-9425-3

Tan. L-H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. *Reading and Writing: An Interdisciplinary Journal, 10,* 165–200

Taouk, M., & Coltheart, M. (2004). The cognitive processes involved in learning to read in Arabic. *Reading and Writing: An Interdisciplinary Journal, 17,* 27-57

Thomas, M.S.C., & McClelland, J.L. (2008). Connectionist models of cognition. In R. Sun (Ed). *Cambridge handbook of computational psychology.* Cambridge University Press, 23-58.

Tolchinsky, L., Levin, I., Aram, D., & McBride-Chang, C. (2012) Building Literacy in Alphabetic, Abjad and Morphosyllabic Systems. *Reading and Writing: An Interdisciplinary Journal*, *25*(7), 1573-1598.

Vaknin-Nusbaum, V., & Miller, P. (2011) The importance of vowel diacritics for the temporary retention of high and low frequency Hebrew words of varying syllabic length. *Memory & Cognition. 39*(3),516-26.

Velan, H., Frost, R., Deutsch, A., & Plaut, D. C. (2005). The processing of root morphemes in Hebrew: Contrasting localist and distributed accounts. *Language and Cognitive Processes*, *20*, 169-206.

Velan, H., & Frost, R. (2011). Words with and without internal structure: What determines the nature of orthographic and morphological processing? *Cognition 118*, 141–156.

Venezky, R.L. (1970). *The structure of English Orthography*. The Hague, Paris, Mouton

Vergara-Martínez, V & Swaab, T.Y. (2012) Orthographic neighborhood effects as a function of word frequency: An event-related potential study *Psychophysiology*, *49*(9), 1277–1289.

Verhoeven, L., & Perfetti, C. A. (2011). Morphological processing in reading acquisition: A cross-linguistic perspective. *Applied Psycholinguistics*, *32*, 457-466.

Warren, P. (2012) *Introducing Psycholinguistics* Cambridge, Cambridge University Press

Was, C.A. (2011). Inhibition of ambiguous word activation in the context of a working memory task. *International Journal of Linguistics 3*(1), E6 Online at www.macrothink.org/journal/index.php/ijl/article/download/.../pdf_8

Weiss, Y., Tami, K., & Tali, B. (2015). Many ways to read your vowels—Neural processing of diacritics and vowel letters in Hebrew. *NeuroImage*, *121*, 10–19.

Wilson, M., Cuetos, F., Davies, R. A. I., & Burani, C. (2013). Revisiting age-of-acquisition effects in Spanish visual word recognition: the role of item imageability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1842-1859.

Wintner, S. (2009). Language resources for Semitic languages: challenges and solutions. In: Nirenburg, S. (ed.) *Language engineering for lesser-studied languages*. IOS, Amsterdam, 277–290.

Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (2006). Single- versus dual-process models of lexical decision performance: Insights from response time distributional analysis. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 13241344.

Yap, M.J., & Liow, S.J.R. (2016). Processing the written word. In V. Cook, & D. Ryan (Eds.) *Routledge Handbook of the English writing system.* Abingdon: Routledge (pp. 453-470).

Yap, M.J., Sibley, D.E., Balota, D.A., Ratcliff, R., & Rueckl, J. (2015). Responding to non-words in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*(3), 597-613. DOI: 10.1037/xlm0000064. Retrieved 4 November, 2015, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4404174.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979.

Yaseen, Q., & Hmeidi, I. (2014). Extracting the roots of Arabic words without removing affixes. *Journal of Information Science*, 1–10. DOI: 10.1037/xlm0000064. Retrieved 16 May, 2017, from http://www.just.edu.jo/~qmyaseen/Arabic1.pdf

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, 47, 1-29.

Zeigler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving neighborhood conflicts in the lexical decision task. *Cognition, 68,* 53–62.

Ziegler, J.C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory *Psychological Bulletin*, *131*(1), 3–29

Appendix A Extract from list of 1,474 unvowelised letter strings

target	modal	length	frequency
ابتداء	اِبْتِدَاء	6	37.56
اتجاه	اِتِّجَاه	5	93.45
اتساق	اِتِّسَاق	5	0.05
اتصال	اِتِّصَال	5	46.74
اتفاق	اِتِّفَاق	5	380.43
اتفاقية	اِتِّفَاقِيَّة	7	85.23
اجتماعي	اِجْتِمَاعِي	7	21.12
اجتمع	اِجْتَمَعَ	5	31.44
احتباس	اِحْتِبَاس	6	0.23
احترام	اِحْتِرَام	6	51.13
احتفال	اِحْتِفَال	6	32.17
احتكار	اِحْتِكَار	6	0.18
احتمال	اِحْتِمَال	6	78.86
احمرار	اِحْمِرَار	5	0.16
اختار	اِخْتَارَ	5	25.62
اختزال	اِخْتِزَال	6	2.91
اختل	اِخْتَلَّ	5	0.86
اختيار	اِخْتِيَار	6	83.75
ارتفع	اِرْتَفَع	5	43.85
ارتقاء	اِرْتِقَاء	6	0.03

Appendix B Extract from list of 1,352 non-words

word	non-word
انتهك	ابتحك
ابتداء	ابضداء
اتساق	اتحاق
اتفاقية	اتخاقية
اتجاه	اتشاه
اتصال	اتمال
اجتمع	اجتسع
احتكار	احتطار
احتفال	احتغال
امرأة	احرأة
احمرار	احكرار
اختيار	اخميار
اختار	اخنار
ارتفع	ارجتع
ارتقاء	ارشقاء
ازدراء	ازعراء
استيطان	اسبيطان
افتتاح	استتاح
استمرار	استثرار
استقرار	استخرار
استنفار	استخفار
احترام	استرام
استشعر	استشضر
استشرى	استضرى

Appendix C Research ethics forms

LANCASTER UNIVERSITY

Applicant: Ahmed Mohammed A. Alhussein
Supervisor: Dr Robert Davies
Department: Psychology

28 May 2014

Dear Ahmed and Robert,

Re: The effect of the printed word attributes on Arabic reading

Thank you for submitting your research ethics application for the above project for review by the Department of Psychology Ethics Committee. The application was recommended for approval by the Department Committee, and on behalf of the Chair of the University Research Ethics Committee (UREC), I can confirm that approval has been granted for this research project on condition that your laptop will be encrypted as well as password protected. (This is required because identifiable data, which includes recordings of participants, must be encrypted in order to comply with Data Protection policy). Further information about encryption can be found on the ISS webpage http://www.lancs.ac.uk/iss/security/guidance/encryption.htm and additional information about this can be requested from the ISS helpdesk.

As principal investigator your responsibilities include:

- ensuring that (where applicable) all the necessary legal and regulatory requirements in order to conduct the research are met, and the necessary licenses and approvals have been obtained;
- reporting any ethics-related issues that occur during the course of the research or arising from the research to the Research Ethics Officer (e.g. unforeseen ethical issues, complaints about the conduct of the research, adverse reactions such as extreme distress);
- submitting details of proposed substantive amendments to the protocol to the Research Ethics Officer for approval.

Please contact the Research Ethics Officer, Debbie Knight (01542 592605 ethics@lancaster.ac.uk) if you have any queries or require further information.

Yours sincerely,

S.C. Taylor

Sarah Taylor
Secretary, University Research Ethics Committee

Cc Fiona Aiken, University Secretary, (Chair, UREC); Professor Gert Westermann (Chair, Psychology DEC)

Research Support Office
Research and Enterprise Services
Lancaster University
Bowland Main
Lancaster LA1 4YT
United Kingdom

Tel: +44 (0) 1524 592002
Fax: +44 (0) 1524 593229
Web: http://www.lancs.ac.uk

بسم الله الرحمن الرحيم

KINGDOM OF SAUDI ARABIA
Ministry of Higher Education
Al-Imam Muhammad Ibn Saud Islamic University
COLLEGE OF SOCIAL SCIENCES
Department Of Psychology

المملكة العربية السعودية
وزارة التعليم العالي
جامعة الإمام محمد بن سعود الإسلامية
كلية العلوم الاجتماعية
قسم علم النفس

الرقم : .... التاريخ : / / ١٤٢ هـ المشفوعات : ....

To whom it may concern

The research title is: "The effect of printed word attribute on Arabic reading"

The Department of Psychology at Imam Muhammad Ibn Saud University confirms that Mr Ahmed Alhussein is a member of the Department's academic staff and he is being funded by the University to do his PhD degree at Lancaster University.

The Department would like to advise you that Mr Alhussein has a full ethics approval to conduct his experiments at IMISIU and Mr Alhussein will receive a full support (e.g., a quiet room, Psychology laboratory).

Dr Saleh Alorini

25-9-2.13

Head of Department of Psychology

الرياض ١١٤٧١ - ص . ب : ٣١٦٩ - هاتف وفاكس رئيس القسم: ٢٥٨٥٩٢٤ - وكيل القسم: ٢٥٨٥٩٢٥ - القسم ٢٥٨٥٩٢١ - ٢٥٨٥٩٢٢
Riyadh 11471 - P.O box: 3169 - Tel & Fax Head Dept. 2585924 - Deputy Dept. 2585925 - Dept. 2585921 - 2585922
E-Mail: Library @ imamu.edu.sa