Topics in Arabic Auditory Word Recognition: Effects of Morphology and Diglossia

by

Moh'd A. Al-Omari

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Linguistics

University of Manitoba

Winnipeg

Abstract

This dissertation investigates the cognitive relevance of Arabic morphology and diglossia in spoken word recognition. The current study asks four main questions: (1) Does Arabic morphology influence word recognition? (2) Which view of Arabic morphology (i.e., the root-based or the stem-based) has an online role in spoken word recognition? (3) Does Arabic diglossia (i.e., using colloquial Arabic (CA) and Modern Standard Arabic (MSA) as the dominant language of speaking and literacy, respectively) affect spoken word processing? (4) How can Arabic diglossia affect spoken word recognition? Three different lexical decision experiments and one phoneme-monitoring task were designed and conducted on a group of 140 literate native speakers of Jordanian colloquial Arabic (JCA).

In the first experiment, the participant responded to MSA words varied in their surface, root, and stem frequencies. Results revealed that the token frequencies of the three tested units affected the speed of word recognition to the same extent. This suggests that both roots and stems, along with the surface words, are valid units of Arabic mental lexicon. The next two experiments compared the processing of JCA and MSA words when embedded in sentences of the same or the other variety of Arabic and when primed by intra-variety vs. cross-variety words. Results showed a lexical switching cost only when the target word is processed in the sentential context. Moreover, while the sentence experiment reported a processing advantage for MSA words relative to JCA words, the priming experiment found a processing advantage for JCA words. The priming effects were larger when the related primes were presented in JCA relative to the priming effects of the MSA primes.

The fourth experiment compared phoneme monitoring of consonants and short vowels in JCA and MSA words. Results showed a detection advantage for consonants relative to short vowels and no difference between the carrier words of the two varieties of Arabic. On the whole, the last three experiments suggest that both spoken language (i.e., CA) experience and literary language (i.e., MSA) experience can affect auditory word recognition. This work emphasizes the relevance of (alphabetic) literacy and experimental task in speech processing.

Acknowledgements

This dissertation is not the outcome of one person alone. Many people helped in this research, and without their contributions this work could not be possible. I greatly appreciate the following people for their contributions.

First and foremost, I would like to express my deep gratitude and indebtedness to my advisor Dr. Kevin Russell for his guidance and insights. Prof. Russell broadened my horizons and opened my eyes to new approaches in the fields of psycholinguistics and statistics. Definitely, without his enthusiasm and patience this research could not see the light. I would also like to thank my committee members, Dr. Robert Hagiwara and Dr. Melanie Soderstrom, for their valuable remarks and advice on this research. I am much obliged to Dr. Laura Sabourin for serving as the external examiner and dedicating considerable time and attention to my dissertation. Special thanks are due to Dr. Erin Wilkinson for supervising my first generals paper. I am also thankful to Dr. Jila Ghomeshi, Dr. Terry Janzen, and Heather Lee, the friendly department administrator.

I am grateful to Dr. Osama Al-Omari and Dr. Ahmad Al-Omari from Yarmouk University for their help in providing subjects of this investigation. I am also very thankful to the staff in the Department of Modern Languages at Yarmouk University for allowing me to use the department facilities and interview my participants. I would like to say thank you to my special friend Mohammad Qassim from the Department of Computer Science at UFM for his assistance in many technical and software programming issues. I would also like to thank Jesse Stewart, Oqab Shawashreh, Mutasim Al-Deaibis, and Meera Sahawneh for their friendship.

None of this would have been possible without the priceless support and sacrifice from my wife, Mrs. Rana Baaj. Finally, I would like to thank my parents and family for years of continuous support.

Dedication

To the five candles lighting my life this dissertation is dedicated.

To my warm and caring parents

To my lovely wife

To my adorable sons, Eyad and Ahmad

Table of Contents

List of Tables

List of Figures

General Introduction

Arabic is a Semitic language, the fifth most widely used languages in the world. Arabic is the official language for 27 states located in Asia and North Africa, and the native language of approximately 300 million people worldwide (Saiegh-Haddad & Henkin-Roitfarb, 2014). Arabic speaking communities and schools also exist in Europe, North America and Australia, as a result of immigration. In Israel, Arabic is an official language along with Hebrew, where approximately 20% of the population use Arabic as their native language (Statistical Central Bureau, 2005, as cited in Levin, Saiegh-Haddad, Hende, & Ziv, 2008).

A common controversy about the Semitic languages, including Arabic and Hebrew, is the structure of their morphology. A group of researchers argue that Arabic and Hebrew have nonconcatenative morphology (e.g., Bauer, 2004; Cantineau, 1950a, 1950b; Glinert, 1996; McCarthy, 1981; Rajhi, 1976; Wright, 1967). According to this perspective of word formation, most Arabic words (e.g., *laaʕib* ‘player’) are nonlinearly and discontinuously made up of two abstract representations: tri-consonantal roots (e.g., {lʕb} ‘related to playing’) and word patterns (e.g., *CaaCiC/agentive nouns*). Consonantal roots convey lexical meanings while word patterns denote morpho-syntactic information. Arabic orthography associates consonantal roots with full letters while it partially encodes word-patterns; short vowels of word patterns are usually left unspecified, except in some children’s books and religious scripts where print encodes them with small diacritic markers (Abu Rabia, 2001). An alternative approach of Arabic morphology proposes that Arabic, like English, is a stem-based language (Benmamoun, 1999, 2003;

Ratcliffe, 1998, 2004). Based on this view, the Arabic words *tˤuwliy/yatˤuwl* 'length/prolong' concatenatively (i.e., linearly) decompose into the suffix *–iy/*the prefix *ya-* and the nominal stem [tˤuwl] rather than into the tri-consonantal root {tˤwl} and the word patterns *CuCCy/yaCCuC*. Benmamoun (1999, 2003) proposes the Arabic imperfect stem (e.g., [ʕallim]) as the base for the formation of some inflected words (e.g., *yu-ʕallim* 'third person masculine prefix+teach') and derived words (e.g., *mu-ʕallim* 'teacher').[1]

Another thought-provoking property of Arabic is that native speakers of Arabic grow up in a special linguistic situation called 'diglossia' (Ferguson, 1959). Speakers of Arabic use two co-existing varieties of Arabic, a colloquial variety and Modern Standard Arabic, in different contexts. While all native speakers of Arabic use the same Modern Standard Arabic (MSA), their spoken dialects vary from one state to another and even in different regions of the same state. MSA is nobody's native language, and Arabs learn it in school as the language of literacy (reading and writing). MSA is used in formal situations, such as public speaking, religious contexts, media, and the press. In contrast, native speakers of Arabic acquire their colloquial Arabic (CA) early and use it in everyday situations when they communicate with each other (Abu-Rabia, 2000; Ferguson, 1959; Maamouri, 1998).

The present research will address the morphological and diglossic characteristics of Arabic from a psycholinguistic point of view. The first goal of my research is to learn which of the two proposed morphologies of Arabic (concatentative vs. nonconcatenative morphology) has a psychological reality in the minds of native Arabic speakers. The

---

[1] See Ussishkin (199, 2005) for stem-based views of morphology in Hebrew.

second goal of this dissertation is to determine whether and in what way diglossia can affect the lexical access of Arabic speakers in behavioral tasks of word recognition. This dissertation consists of four self-contained experiments. The first experiment is concerned with Arabic morphology while the other three studies are related to Arabic diglossia. The morphology experiment is not very linked to the diglossia experiments. This is because there is no database available for (Jordanian) colloquial Arabic. When such a database is available, it becomes easier to investigate whether literacy and diglossia affect results of the first experiment, or how morphology may have an effect on the data from the diglossia experiments.

Chapter One

Study I: The Role of Arabic Morphology in Auditory Word Recognition

1.1 Introduction

Morphology is the study of internal word structure and formation, and morphemes are the basic units of word structures. Morphemes form bundles of orthographic, phonological, semantic, and grammatical information that tends to be constant across different but related words (Katamba & Stonham, 2006). Morphemes are economical building blocks of novel words and give grammatical and derivational meanings to existing words. Complex words consist of lexical, derivational, and/or inflectional morphemes. Lexical morphemes encode the content meaning of the word, derivational morphemes change the meaning of the word or its part of speech (e.g., verbs into nouns), and inflectional morphemes add extra grammatical information to the word (e.g., person, number, gender). For example, the English complex word *schools* can be decomposed into the free stem *school* (a lexical morpheme) and the plural suffix -*s* (an inflectional morpheme). Similarly, the English complex word *player* is derived from the stem *play* and the agentive suffix -*er* (i.e., a derivational morpheme).

Morphemes have attracted researchers' attention as potential constituents that promote efficiency in word storage, retrieval and processing. Over the last three decades, behavioral research on morphology and word recognition has focused on two core questions (for some review, see Diependaele, Grainger, & Sandra, 2012). The first is, "Does morphological knowledge affect word recognition?" The answer to this question demonstrates whether morphological constituents (i.e., stems, suffixes, prefixes, infixes) per se influence word processing independent of other connected factors such as form (phonological/orthographic) or semantics (meaning). The second question is “What is the mechanism that causes this possible effect of morphology? This latter question is more debatable, as it argues about the locus of morphological representations in the lexicon.

1.1.1 Approaches of Arabic Morphology

This research focuses on two different views of Arabic morphology: the root-based morphology and the stem-based morphology.[2] Neither of these views is homogeneous as, there are many perspectives within each approach (Boudelaa, 2014). Researchers refer to three root-based approaches of Arabic morphology. For medieval Arab grammarians, Arabic consists of two basic morpheme types: a consonantal root, which conveys the lexical (i.e. referential) meaning of the word, and the vocalic word pattern, which forms the Arabic *masˤdar* (i.e., the deverbal noun stem). All other Arabic words do not involve a root and word pattern combination. Instead, surface forms of these words are derived from the deverbal noun stem by morpho-phonological processes, such as affixation, and vowel insertion and deletion. On this view of root-based morphology, the tri-consonantal root {slb} ‘rob’ is initially combined with the *CaCC* template to form the deverbial noun stem [salb] ‘robbing’. Then the perfect verb [salab] ‘robbed’ is built by inserting the short vowel [a] before the final consonant of the source noun [salb]. The passive participle *masluub* ‘robbed’ is also derived from the same deverbal source noun by adding the prefix [ma-], removing the medial vowel and inserting the long back vowel [uu] before the final consonant. Another group of medieval grammarians, followers of the *Kufa School* of grammar, proposed the perfect form of a verb rather than the source noun as the starting point for all other derivations.

---

[2] Research on Indo-European word recognition uses the terms stem and root interchangeably to refer to the same thing. In Semitic research, however, the two terms refer to two different types of morphology.

In contrast to this view, a group of structuralist-linguists argue that every surface form of Arabic words is the outcome of combining a consonantal root with some morpho-syntactic pattern (e.g., Cantineau, 1950a, 1950b; Cohen, 1951; Hilaal, 1990). For example, the surface forms *salb* ‘robbing’, *salaba* ‘robbed’, and *masluub* ‘robbed’ share one underlying root {slb}, but are mapped onto three different word patterns (i.e., templates): *CaCC* {faʕl} ‘deverbal noun or gerund’, *CaCaCa* {faʕala} ‘perfective verb’ and *maCCuuC* {mafʕuul} ‘passive participle’, respectively.[3] The third view of the root-based approach is developed within autosegmental phonology (McCarthy 1979, 1981, 1982). Proponents of this perspective agree with some structuralists’ view that Arabic words consist of consonantal roots. Alternatively, they suggest two tiers of morpho-syntactic information instead of the template. The first tier is the vocalic melody, which conveys grammatical information such as passive-active, whereas the second tier represents the CV-skeleton structures, which convey rich varieties of morpho-syntactic information (e.g., perfectiveness, reciprocal meaning, and locative meaning). For instance, the surface form *salab* ‘robbed’ is built up of the tri-consonantal root {slb} ‘rob’, the *a-a* vocalic melody, which indicates declarative meaning, and the CVCVC structure, which indicates the perfective tense of the verb.

Like root-based morphology, Arabic stem-based morphology is not uniform and has different perspectives. Heath (2003) argues that word patterns do not contribute to any grammatical information in many cases. For example, the *CuCC*, *CaCC*, and *CiCC* patterns do not convey any grammatical information in words such as *xubz* ‘bread’, *ħaql* ‘field’ and *ħilm* ‘forgiveness’. In this view, Arabic has a core of basic non-derived stems,

---

[3] The {f ʕ l} in any word template (e.g., {faʕl}, {mafʕuul}, etc.) are slots that can be replaced by any tri-consonantal root.

such as singular nouns and imperfective of simple verbs. These non-derived stems can contribute to the derivational processes by ablaut (i.e., vowel change) and/or affixation. Ratcliffe (2004) reports a different view of Arabic stems. He suggests a sonority-based mechanism to extract word affixes and recover the stems. In a third view, Benmamoun (1999, 2003) introduces the indicative imperfective verb as a plausible candidate for some word formation. Benmamoun demonstrates that the base forms of imperfect verbs (e.g., *yu-**ʕallim***) '3.singular-teach/he teaches' take the central role in the formation of Arabic nominals (e.g., *mu-ʕallim*) 'teacher', imperatives (e.g., *ʕallim*) 'teach', and locatives (e.g., *ya-**sbaħ**/ma-sbaħ*) '3.singular-swim/swimming pool'. Benmamoun prefers imperfective to perfective verbs as central stems of Arabic because words derived from imperfective verbs keep stem vowels intact. Alternatively, using perfective verbs as input forms involves extra procedures of vowel changes (e.g., *ʕallama/muʕallim* or *sabaħa/masbaħ*).

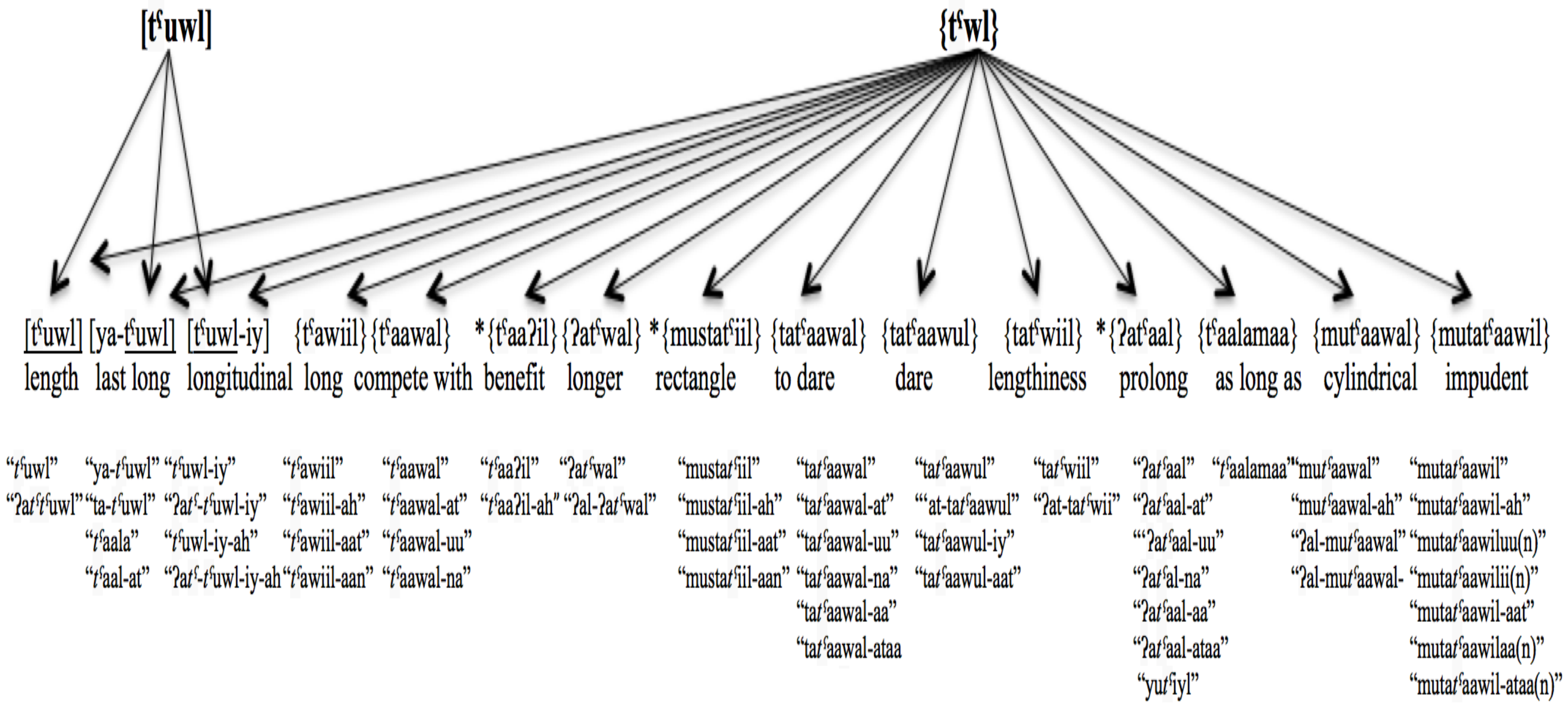
The study presented in this chapter does not intend to seek support for one perspective of the root-based and/or stem-based approaches over any other approach. No matter which of the two root-based versions (i.e., McCarthy or some structuralists' approach) is to be adopted, both of them predict the cognitive relevance of the Arabic root, the constituent to be tested against the stem in this work. This study will also abstain from committing to any of the above views of Arabic stems, though it is relatively closer to Benmamoun's perspective. The present study defines Arabic stems in terms of what is most likely to be relevant in auditory word recognition. For the purpose of this research, stems cover any possible linear string of sounds that is recurrent and meaningful, irrespective of its word class. For example, the verb [ʕaskar] 'to camp' is a possible concatenative stem since it recurs in many semantically-related words such as *ta-ʕaskar* ‘to gather in a place’, *mu-ʕaskar* ‘a camp’, *ʕaskar-ah* ‘camping’, *ʕaskar-iyy* ‘military’, and *ʕaskar-iyyah* ‘military service’. A stem can also be a noun such as *ʃamʕ* ‘wax’, which appears in the related words *ʃamʕ-ah* ‘candle’ and *ʃamʕ-iyy* ‘waxy’. These constituents are attached to what could be real prefixes and suffixes. This definition of Arabic stems helps us employ morphological families of different sizes and frequencies (see Section 1.1.2). It also draws a clear boundary between the putative linear and non-linear morphologies of Arabic.

1.1.2 Purpose of the Study

The first part of my dissertation will examine the role of three frequency measures in Arabic lexical access (i.e., processing spoken Arabic words). These frequencies are the *whole-word frequency* (i.e., the number of times a word occurs in a corpus), the *cumulative root/stem frequency* (i.e., the token-based frequencies of all words that share the same stem/root) and the *morphological family size* (i.e., the number of different derived words that contain the same stem/root, excluding inflectional words). The three frequency counts are usually linked with particular models of morphological processing (Ford, Marslen-Wilson, & Davis, 2003), and used to inform different theories of how morphologically complex words are represented and processed in our lexicon (i.e., mental dictionary; Diependaele et al., 2012). The present study will examine the possible effects of token and type counts of Arabic morphemes with special reference to the proposed concatenative and nonconcatenative morphologies of Arabic. It will scrutinize which frequency of the possible morphological constituents of Arabic, stem-based frequency or root-based frequency, has an online role in the auditory word recognition of

Arabic. It is possible that different gauges of Arabic morphological frequencies (stem-based vs. root-based frequency) have different effects on how fast native speakers of Arabic access Standard Arabic words.

Figure 1.1 visualizes the morphological paradigm for the Arabic word *tˤuwl* 'length'. The surface frequency of *tˤuwl* is the number of times *tˤuwl* appears in the corpus: *F (tˤuwl)*. The stem family-size of *tˤuwl* is the number of derived words that contain the same sequential stem [tˤuwl]: *NS (tˤuwl) = tˤuwl, ya-tˤuwl, tˤuwl-iy* = three tokens. The root family-size of the Arabic word *tˤuwl* is the number of derived words that share its tri-consonant root {tˤwl}; *NR (tˤuwl) = tˤuwl, ya-tˤuwl, tˤuwl-iy, tˤawiil, tˤaawal, tˤaaʔil, ʔatˤwal, mustatˤiyl, tatˤaawal, tatˤaawul, tatˤwiil, ʔatˤaal, tˤaalamaa, mutˤaawal, mutatˤaawil* = 15 tokens. The cumulative stem frequency for *tˤuwl* is the total sum of inflectional frequencies for the three derived words *tˤuwl*, *yatˤuwl* and *tˤuwliy*: *FS (tˤuwl)*= *F* [tˤuwl] + *F* [yatˤuwl] + *F* [tˤuwliy] **=** *F (tˤuwl)* + *F (ʔatˤtˤuwl)* + *F (yatˤuwl)* + *F (tatˤuwl)* + *F (tˤuwliy)* + *F (ʔatˤtˤuwliy)* + *F (tˤuwliyah)* + *F (ʔatˤtˤuwliyah)*. Finally, the cumulative root frequency for *tˤuwl* is the total number of inflectional frequencies of all words that share its root including the words, in the small print, inflected for person, number, gender and definiteness: *FR (tˤuwl)* = *F* {tˤuwl} + *F* {yatˤuwl} + *F* {tˤuwl-iy} + *F* {tˤawiil} + *F* {tˤaawal} + *F* {tˤaaʔil} + *F* {ʔatˤwal} + *F* {mustatˤiyl} + *F* {tatˤaawal} + *F* {tatˤaawul} + *F* {tatˤwiil} + *F* {ʔatˤaal} + *F* {tˤaalamaa} + *F* {mutˤaawal} + *F* {mutatˤaawil} **=** *F (tˤuwl)* + *F (ʔatˤtˤuwl)* **…** + *F (mutatˤaawil-ataa)*, etc.



* The glide in the tri-consonantal root {*t*ˤwl} changes into [aa] and [ii] in some words.

Figure 1.1: *Morphological paradigm for the word "tˤuwliy" (linear and nonlinear analysis)*

Each of these possible morphological accounts of Arabic has some advantage over the other. The stem-based account might be a stronger predictor for Arabic word recognition, relative to the root-based account, by virtue of the intense phonological overlap between words of the same stem (i.e., qualitative advantage). According to this argument, when effects of other predictors are held constant, how fast speakers of Arabic react to a word such as *tˤuwliy* is (more) related to the frequency of its stem [tˤuwl] rather than the frequency of its tri-consonantal root {tˤwl}. This is despite the fact that the morphological frequencies of a consonantal root, both type-based and token-based, are larger than the morphological frequencies of its stem (remember: the stem-based family is a subset of the root-based family). It is also possible that the root-based account is the only/more dominant predictor in spite of the partial phonological (form) overlap between words and their shared consonantal root, as roots link more Arabic words than stems do (i.e., quantitative advantage).

This study will help us better understand which morphological account of Arabic is a (more) valid predictor of Arabic word processing (i.e., the stem-based account vs. the root-based account). To achieve this target, this experiment will manipulate some Arabic words that vary in their stem and root frequencies. The current study will examine the frequency of Arabic roots and stems rather than word patterns and affixes. This is because previous research has shown marginal effects for Arabic word patterns (e.g., Boudelaa & Marslen-Wilson; 2011) and Indio-European affixes (e.g., Burani & Thornton, 2003), compared to the content morphemes of the tested words.

1.1.3 Behavior Tasks of Word Recognition

Empirical behavioral research has employed different metalinguistic techniques to investigate the relationship between morphology and word recognition. Morphological priming (e.g., Murrell & Morton, 1974) and morphological frequency manipulations (e.g., Baayen, Dijkstra, & Schreuder, 1997) are the most popular techniques. The priming technique investigates the effects of a previously presented stimulus (the prime) on a subsequent stimulus (the target). A positive priming effect appears when participants respond faster to a target word preceded by a (morphologically-) related word compared to the same target word preceded by an (morphologically) unrelated word (e.g., Reid & Marslen-Wilson, 2003; Stanners, Neiser, & Painton, 1979). There are three common priming techniques used in word recognition: simple priming, masked priming, and cross-modal priming. In simple priming manipulation, the prime is presented overtly either visually or aurally before a written/spoken target (e.g., Drews & Zwitserlood, 1995; Murrell & Morton, 1974; Reid & Marslen-Wilson, 2003). In masked priming, the prime is visually introduced for a short period of time and the participant is almost unconscious of its presence (e.g., Deutsch, Frost, & Forster, 1998; Diependaele, Sandra, & Grainger, 2005; Forster & Davis, 1984). In cross-modal priming, primes and targets are presented in two different modalities: auditory prime and visual target or visual prime and auditory target (Boudelaa & Marslen-Wilson, 2004 b; 2004 b; 2011; Marslen-Wilson et al., 1994; Meunier & Segui, 2002).

Morphological frequency is another method adopted to investigate the effects of morphological constituents (i.e., morphemes) on the lexical decision. In this technique, participants are expected to react to a single (unprimed) word based on its overall morpheme frequencies. This method was used in both visual modality (e.g., Neijt et al., 2003; Schreuder & Baayen, 1997) and auditory modality (e.g., Balling & Baayen, 2008; Wurm et al., 2006) to evaluate various theories of simple and complex word processing and representation.[4] In the present study, I will use the single lexical-decision task of word recognition after gauging the different frequency counts of the tested words.

1.1.4 Token Frequency, Type Frequency and Word Recognition

A simple measure of the total number of tokens of a word in a given corpus, surface frequency is one of the earliest and most common findings in lexical decision tasks (e.g., Broadbent, 1967; Howes, 1957). Research found that subjects recognize high-frequency words faster than low-frequency words (i.e., they show shorter response latencies to the more frequent words in lexical decision tasks). The positive effects of surface frequency were shown in both visual modality (e.g., Baayen et al., 2007; Burani & Caramazza, 1987; Taft, 1979) and auditory modality (e.g., Baayen et al., 2007; Meunier & Segui, 1999; Wurm et al., 2006). The effect of surface frequency was also reported across different languages: English (e.g., Niswander, Pollatsek, & Rayner, 2000; Taft, 1979; Wurm, 1997), French (e.g., Colé, Beauvillain, & Segui, 1989; Meunier & Segui, 1999), and Italian (e.g., Burani & Caramazza, 1987).

The effects of type-count frequency (family size) also prove valid on both morphologically simple words (Baayen, Tweedie, & Schreuder, 2002; De Jong,

---

[4] For more visual studies, see Burani and Thornton (2003); De Jong, Feldman, Schreuder, Pastizzo, and Baayen (2002); Schreuder and Baayen (1997).

For more auditory studies, see Baayen, McQueen, Dijkstra, and Schreuder (2003); Baayen, Wurm, and Aycock (2007); Meunier and Segui (1999).

Schreuder, & Baayen, 2000; Schreuder & Baayen, 1997) and morphologically complex words (e.g., Bertram, Schreuder, & Baayen, 2000; Ford et al., 2003; Traficante & Burani, 2003).[5] A growing body of research reports positive effects for the type-count frequency in the visual domain of word recognition (e.g., Ford et al., 2003; Neijt et al., 2003) and the auditory modality (e.g., Balling & Baayen, 2008; Wurm et al., 2006). The family size effect was established in different languages such as English (e.g., Feldman & Pastizzo, 2003; Pylkkänen, Feintuch, Hopkins, & Marantz, 2004), Dutch (e.g., De Jong et al., 2000; Dijkstra, Moscoso del Prado Martin, Schulpen, Schreuder, & Baayen, 2005; Schreuder & Baayen, 1997), German (Lüdeling & De Jong, 2002), Danish (Balling & Baayen, 2008), Hebrew (Moscoso del Prado Martin et al., 2005), and Arabic (Boudelaa & Marslen-Wilson, 2011). In their fourth experiment of auditory lexical decision, Meunier and Segui (1999) reported another relevant type of frequency. The authors demonstrate that lexical decision times for suffixed French words depend on the number of morphological cohorts that are more frequent than the tested word, given that their other frequencies are controlled. For instance, the French word *crachement* was reacted to faster than *griffement*. While the two words have comparable surface and stem frequencies, *crachement* has fewer more frequent candidates in morphological cohorts (5 morphological competitors) relative to *griffement* (13 morphological competitors). This direct correlation between reaction times and the number of competing candidates in the morphological cohort needs further investigation. Hereafter, I will refer to this type of frequency as numbers of morphological competitors.

---

[5] For further research on complex words, see De Jong et al. (2000); Moscoso del Prado Martin, Bertram, Haikio, Schreuder, and Baayen (2004); Neijt, Schreuder, and Baayen (2003).

The token frequency has less decisive and unambiguous effects compared to the surface and family size counts. Some studies from both visual and spoken modality reported facilitative (i.e., positive) effects for morphological token frequency in various languages: English (e.g., Niswander et al., 2000; Taft, 1979), Dutch (Schreuder, Burani, & Baayen, 2003), French (e.g., Beauvillain, 1996; Colé et al., 1989; Holmes & O'Regan, 1992; Meunier & Segui, 1999), and Italian (e.g., Burani & Caramazza, 1987; Burani & Thornton, 2003). Other research demonstrated inhibitory (i.e., negative) effects for the token frequency (e.g., Baayen et al., 2002; Ford et al., 2003). However, some experiments found no effects at all when the stem-type counts were controlled or partialled out in both visual modalities (e.g., Bertram et al., 2000; Sereno & Jongman, 1997) and spoken modalities (Baayen et al., 2007; Balling & Baayen, 2008; Wurm, Ernestus, Schreuder, & Baayen, 2006).[6] Different models of word recognition suggest different architectures for the lexical representations of words and predict different effects for the whole-word and morpheme frequencies. The next section introduces a brief account of these models and what they hypothesize about word and phoneme frequencies.

1.1.5 Frequency and Models of Lexical Representation

Although the representational locus of word and morpheme frequencies is still debatable and unclear (for details see Diependaele et al., 2012; Ford et al., 2003), in what follows, I will summarize the most dominant views of how these frequencies are relevant to word

---

[6] As word recognition is a complex process with many predictors involved, researchers find it difficult to control all possible predictors. Accordingly, they usually consider and look beyond the other common predictors in their statistical models when they test the effects of the new variable(s) under investigation. This approach helps the researcher to partial out (i.e., eliminate) the effects of the formerly attested predictors.

recognition. The surface frequency effect is consistent with the *full-listing model* of word recognition. According to the *full-listing model*, a word is processed, represented, and accessed as one whole unit in the lexicon. Sereno and Jongman (1997) propose a strong version of the *full-listing hypothesis*: complex and simplex words are only represented as fully formed in the mental lexicon. According to the strong version of the *full-listing hypothesis*, it is the surface frequency rather than the morpheme frequency of the word that affects its access in the mental lexicon (Bybee, 1995; Lukatela, 1991). Katz, Rexer, and Lukatela (1991) state that morphological parsing is necessary only when readers read new words. The *full-decomposition model* is an opposing model (Bresnan, 1982; Lieber, 1992; Stockall & Marantz, 2006). According to the pure morpheme listing view, the only way sublexical information (e.g., phonemes, letters) of words is mapped onto meaning is via morphemes. If this is the case, then only morpheme frequencies matter in word recognition. This approach may look attractive from the economical perspective of language (i.e., storing one morpheme instead of several separate words). Nonetheless, word meanings are not always predictable from the meaning of their morphological components (e.g., depart/department).

The third camp of word recognition argues for the importance of both word and morpheme frequencies in lexical processing (e.g., Baayen et al., 2007; Balling & Baayen, 2008; Taft, 1979; Meunier & Segui, 1999). The point of debate among proponents of this mixed approach of word processing is how word and morphemic constituents of language are accessed. The sublexical view of the mixed approach proposes that morphemic information is used to access whole-word representation. This approach is also known as the *obligatory decomposition model* (Taft, 1985; 2004; Taft & Forster, 1975), or the *direct access model* (Marslen-Wilson, Tyler, Waksler, & Older, 1994; Marslen-Wilson & Zhou, 1999). The sublexical model assumes that morphemes are the first and default lexical representations used to access whole words, which, in turn, come at a later stage after morphological parsing. Some proponents of the sublexical approach give evidence for the decomposition of both suffixed and prefixed words. For instance, Taft (1979) conducted visual lexical-decision experiments and found that cumulative stem frequencies of both English prefixed and suffixed words affect lexical decision times. In a more recent study, Kazanina (2011) found a priming effect between simple Russian primes and their prefixed, morphologically-related targets during visual perception.

Evidence for prefix-stem parsing can also be extended to auditory word recognition (e.g., Taft, Hambly, & Kinoshita, 1986; Wurm, 1997; Wurm & Aycock, 2003; Wurm et al., 2006). For example, Wurm (1997) demonstrates that morphological decomposition can happen in prefixed words that are semantically transparent and have highly recurrent prefixes. Prefixed words with high morphological family size also undergo morpheme stripping (Wurm et al., 2006). Wurm (and the coauthors) emphasize that prefixed words do not only decompose, but also decompose at an early stage of processing. The author(s) examined the relevance of two types of uniqueness point in prefixed word processing: the traditional *uniqueness point* (UP) and the *conditional root uniqueness point* (CRUP). UP is the point at which there is only one possible candidate left in the cohort of the other competing words (Marslen-Wilson, 1987), (e.g., the second /d/ in /diskred̲ət/). CRUP, on the other hand, is the uniqueness point of the free stem of a prefixed word, given the prefix in question (Wurm & Aycock, 2003), (e.g., the /r/ sound in /diskr̲edət/). CRUPs are different from UPs in many prefixed words. The authors found significant effects for both types of uniqueness points, and introduced the relevance of CRUPs as evidence for an early prefix stripping procedure.

Other researchers argue that word decomposition occurs in suffixed but not prefixed words (e.g., Meunier & Segui, 1999; 2003; Schriefers, Zwitserlood, & Roelofs, 1991; Tyler, Marslen-Wilson, Rentoul, & Hanney, 1988;). This is because morpheme parsing operates from left to right, and for prefixed words stem processing does not precede whole-word processing (Meunier & Segui, 2003). For example, Meunier and Segui (2003) propose that only cumulative frequencies of spoken suffixed words can be relevant in spoken word recognition. The authors compared lexical decision times for some free stems of French and prefixed words derived from these stems. The authors measured participants' reaction times from the beginning of the stem in both types of words after making the length of both stems constant. Meunier and Segui found that free stems elicit longer reaction times than prefixed stems do. Their results are consistent with the hypothesis that identifying prefixed words is not delayed until recognizing their stems. Similar results can also be found in the visual modality. Colé et al. (1989) reported cumulative stem frequency effects for written suffixed words, where the stem is processed first, but not for written prefixed words. All in all, proponents of the sublexical approach concur with the view that polymorphemic suffixed words involve morphological decomposition prior to identifying the whole word. What those researchers disagree on is whether stem decomposition is applicable to (early) prefixed word identification as well.

A second mixed model suggests a supralexical view of polymorphemic words (i.e., the surface word is processed before its morphological components). According to this model, stem/root morphemes are accessed at the lexical level, and morphological representations are superimposed on the whole-word architecture, (e.g., Frost, Forster, & Deutsch, 1997; Giraudo & Grainger, 2000; Grainger, Colé, & Segui, 1991). For example, Giraudo and Grainger (2000) propose that morphemic representations can provide positive feedback to all matching whole-word representations.

The *morphological race model* (Baayen et al., 1997; Schreuder & Baayen, 1995) is a third possible mixed model that looks at complex words as both full word (i.e., the full listing) and separate morpheme (i.e., morphological compositionality) units of the lexicon. According to the *morphological race model*, both surface word and morpheme frequencies are valid predictors for word recognition because linguistic inputs are mapped onto corresponding whole-word and morphemic representations in parallel dual-route architecture. That is, the full form and the morpheme constituents are processed simultaneously and race against each other. Whether the morphological or the full-form processing wins the competition depends on variables such as their frequencies: highly-frequent complex words will be processed as whole words (the direct route) and show the full-form frequency effect, whereas low-frequency complex words will be decomposed into their morphological components (indirect route) showing a morpheme-frequency effect (Alegre & Gorden, 1999; Baayen et al., 1997; Balling & Baayen, 2008; Schreuder & Baayen, 1995). The sublexical, superlexical and race models of morphology reviewed so far are hybrid models, focusing on how meaningful constituents (i.e., words and morphemes) of language are stored and accessed in the mental lexicon. The remaining part of this section will briefly describe how a distributed model conceives of word processing.

The *distributed connectionist model* of language processing (e.g., Plaut & Gonnerman, 2000) is another approach proposed to be inherently sensitive to whole words and morpheme frequencies (Seidenberg & McClelland, 1989). According to the distributed connectionist accounts of word recognition, words are not single nodes but the co-activation of distributed phonological, orthographic and semantic features. For example, the word dog is the co-activation of its orthographic, phonological, and semantic features (e.g., can bark). This co-activation emerges through hidden units, which are the outcome of routine and trained mapping between (sublexical) form and meaning. The *distributed connectionist model* proposes that activating related words is a continuous gradable issue rather than an all-or-nothing phenomenon (Hay & Baayen, 2005; Rueckl & Raveh, 1999). How robust the network of related words is activated depends on factors such as their formal similarity, their semantic consistency (Gonnerman, Seidenberg, & Andersen, 2007; Plaut & Gonnerman, 2000;), and their frequency (e.g., Davis, van Casteren, & Marslen-Wilson, 2003). The *distributed connectionist model* proposes that both whole-word and morphemic representations originate at the same level(s) and undergo the same process.

The full-listing, only-morphemes, mixed and distributed models of word recognition have differing predictions about which frequency measures are relevant in word processing. While mixed and distributed models predict the relevance of both word and morpheme frequencies, the *full-listing model* and *full-decomposition model* argue for the unique relevance of surface frequency and morpheme frequency, respectively. The present research will attempt to evaluate these models in light of the attained results. The following section will summarize some findings on how native speakers of Arabic process Arabic words.

1.1.6 Arabic Word Recognition

Psycholinguistic research presents some evidence for roots and word patterns as building blocks of lexical organization and access in Arabic. A number of priming experiments, including masked priming, cross-modal priming, and auditory-auditory priming, found that Arabic words that share the same roots (e.g., *k̲it̲aab̲/k̲aat̲ib̲, ‘book/writer’*) prime (i.e., activate/facilitate) each other (e.g., Boudelaa & Marslen-Wilson, 2000; 2001; 2005; Mahfoudhi, 2007). Priming effects of Arabic roots were reported even when the relationships between the roots were opaque in their semantics (e.g., *kitaab/katiibah,*‘book/battalion’) and/or forms (e.g., *it̲tif̲aaq̲un/w̲aaf̲aq̲a,*‘agreement-agree’) (Boudelaa & Marslen-Wilson, 2004 a; 2004 b; 2005; Mahfoudhi, 2005). Research also found priming effects for word patterns. For example, the prime *sa̲dʒda̲h̲* ‘prostration’ facilitates processing the target word *la̲ʕna̲h̲* ‘curse’ since the prime and the target share the same word pattern *CaCCah*. However, the facilitative effect of the word pattern was not as robust as the facilitative effect of the tri-consonantal roots. In masked priming, primes started to facilitate targets of the same roots at 36 ms stimulus onset asynchrony (i.e. SOA; the delay between the onset of the masked prime and the onset of the target) while primes facilitated targets of the same word patterns at longer SOAs (48 & 64 ms) (Boudelaa & Marslen-Wilson, 2005). The researchers ascribed the priming difference to the fact that roots denote the primary lexical meaning while word patterns carry the secondary morpho-syntactic information (Boudelaa & Marslen-Wilson, 2005). Developmental data from Hebrew supported this account and found that children speakers of Hebrew can manipulate roots at 3 years old while they cannot manipulate Hebrew word patterns before the age of 10 (Ravid & Malenky, 2001). In two auditory-auditory experiments, Boudelaa and Marslen-Wilson (2013) reported that the priming effects occur not only in Modern Standard Arabic but also in dialectal Arabic, where there is no formal teaching of language structure. The authors present this latter finding as evidence against the assumption that the nonconcatentative morphology of Arabic is the outcome of literacy.

Boudelaa and Marslen-Wilson (2011) investigated effects of Arabic root and word-pattern family sizes on word recognition through masked priming and cross-modal (i.e., auditory-visual priming) experiments. The authors manipulated the productivity of both roots and word patterns of the primes. Results showed that word pattern priming was determined by the productivity of the root regardless of the productivity of the word pattern. This indicates that the participants processed the roots earlier than the word patterns. In the context of highly productive roots, word-pattern priming effects happened because the participants quickly accessed the root of the prime and used the remaining time to process the word pattern of the same prime. This, in turn, explains why the productivity of roots is more important than the productivity of word patterns when both primes and targets are being processed. In a recent study, Boudelaa and Marslen-Wilson (2015) reported cross-modal priming effects within nominal and verbal word patterns, and across roots of nouns (e.g., *ʕaql* ‘mind’) and verbs (e.g., *taʕaqqala* ‘be mindful’). Again, the root priming effect was robust and nearly the same for both semantically transparent prime-target words (e.g., *ʕaql* ‘mind’, *taʕaqqala* ‘be mindful’) and semantically opaque prime-target words (*mataaʕ* ‘commodity’, *mutʕah* ‘pleasure’). Some external evidence about the mental representation of Arabic roots and templates comes from slips of the tongue, aphasic errors and language-games (e.g., Idrissi, Prunet, & Beland, 2008; Prunet, Beland, & Idrissi, 2000). All of these results validate roots and templates as abstract lexical entities of Arabic morphology.

Some research on Arabic word processing, however, found no evidence for roots or word patterns as morphological constituents in Standard Arabic (Abu Rabia & Awwad, 2004). Other research found evidence for suffixed-word parsing in Arabic (Mimouni, Kehayia, & Jarema, 1998). Mimouni and her colleagues tested the auditory priming effects of sound plurals (i.e., a stem and plural affix take place: *lbas ‘dress’* vs. *lbas-at ‘dresses’*) and broken plurals (i.e., roots and word patterns: *mətˤrəg məCCəC ‘stick’* vs. *mtˤarəg mCaCəC ‘sticks’*) on normal and aphasic speakers of Algerian Arabic.[7] The findings revealed a facilitative effect for both types of plurals when the primes and the targets were morphologically-related compared to the unrelated priming condition. Furthermore, the broken plurals were accessed faster than the sound plurals. The authors explained that the priming effect between morphologically-related items is “the result of the simultaneous activation of all members of a family when access to one of the members is achieved” (p.79). To account for the second finding, the researchers postulated that broken plurals have fast whole-word access whereas sound plurals decompose into stems and suffixes leading to a slower reaction time. In another behavioral study, Abu Rabia and Awwad (2004) found that Arabic roots and word

---

[7] All the stimuli were in Algerian Arabic.

patterns did not prime related target words in a masked priming task, at a display time of 50 ms, nor did they speed up their recognition in a naming task. Based on these results, they concluded that Arabic words in nominal patterns of derivational morphology have an independent lexical representation. They also suggested that if morphology has any role in Arabic word recognition, it would probably be found in the linear stem-based pattern of Arabic. The results of these studies argue against roots and word patterns as readily accessed constituents of Arabic words.

On the basis of the reviewed literature, we conclude that the validity of roots and word patterns as basic building blocks of Arabic morphology remains inconclusive from a theoretical perspective (e.g., Benmamoun, 1999; McCarthy, 1975; Ratcliffe 1998) and in some experimental research (e.g., Abu Rabia & Awwad, 2004; Boudelaa & Marslen-Wilson, 2004 a; 2004 b; 2011). The purpose of this study is to extend and consolidate the results of earlier research on Arabic morphology and word recognition. It will try to provide an answer to the role of roots, stems, and whole words in Arabic lexical processing. Unlike the other behavioral studies of Arabic word processing, which used priming tasks, the present study will implement a single auditory lexical-decision task where the target words are uncontaminated with any morphologically-related primes but vary in their stem, roots, and whole-word frequencies.

1.2 Research Questions and Hypotheses

The present study will attempt to answer the following questions:

1- Is there a relationship between surface word frequency and how fast speakers of Arabic react to spoken words of Arabic?

2- Is there a relationship between the cumulative root frequency of Arabic words and how fast speakers of Arabic react to these words?

3- Is there a relationship between the root family-size of Arabic words (i.e., type-count frequency) and how fast speakers of Arabic react to these words?

4- Is there a relationship between the number of root morphological competitors for a given Arabic word and its response latency?

5- Is there a relationship between the cumulative stem frequency of Arabic words and how fast speakers of Arabic react to these words?

6- Is there a relationship between the stem family-size (i.e., type-count frequency) of Arabic words and their response latencies?

7- Is there a relationship between the number of stem morphological competitors for a given Arabic word and its response latency?

8- Which approach of Arabic morphology (root-based vs. stem-based) is a better predictor of Arabic lexical access?

Word frequency is one of the earliest variables found affecting the speed of word recognition. Previous research on spoken word recognition reported that the increase in the surface frequency of the word corresponds to an increase in the rate of its recognition (e.g., Cleland, Gaskell, Quinlan, & Tamminen, 2006; Connine, Mullennix, & Yelen, 1990). These findings on the relationship between word frequency and word recognition lead to the following hypothesis:

H1.1: Words are effective units in the lexical access of Arabic: the higher the word frequency is, the easier it is to be accessed.

Moreover, based on the discontinuous approach of Arabic morphology (i.e., Arabic words are derived from consonantal roots and word patterns), this study hypothesizes:

> *H2.1: Roots are effective units in the lexical access of Arabic. Hence, their co-occurrences and productivities are expected to accelerate Arabic word processing.*

The current study will support this hypothesis if it finds enough evidence for a remaining positive relationship between root frequencies of the target words and how fast they are processed after taking into account the effects of all the other competing explanations. Moreover, the proposal that Arabic has a continuous morphology (i.e., Arabic complex words consist of stems, prefixes, and suffixes) leads to the following hypothesis:

> *H3.1: Arabic stems are valid units in the lexical access of Arabic; consequently, their type and token frequencies codetermine how fast words containing these stems are accessed and recognized.*

The investigation will support this hypothesis if speakers of Arabic show a positive correlation between how fast they react to the target words and how frequent and productive their stems are. Again, this possible correlation will be determined after taking into account the effects of all the other competing explanations. A fourth possible hypothesis we can postulate in the context of the current research based on the strong phonological overlap between words that share the same stem is:

H4.1: stem-based frequencies have stronger explanatory power for Arabic word processing than root-based frequencies do.

As Arabic roots are more productive than Arabic stems (i.e., the stem-based frequency of a word is a subset of its superordinate root-based frequency), we predict:

H5.1: root-based frequencies codetermine Arabic word-processing better than stem-based frequencies do.

Hypotheses 4 and 5 suggest that both stems and roots are readily accessed entities of the Arabic lexicon, but they differ in their expectations about which constituent has the primary role and which one has the secondary role in Arabic word recognition.

1.3 Method

1.3.1 Participants

Thirty native speakers of Jordanian Arabic took part in this experiment. The participants were third and fourth-year undergraduate students from Yarmok University, located in northern Jordan. All participants had completed at least 12 years of formal education in Modern Standard Arabic. Most of the participants reported intermediate and high-intermediate knowledge in a second language, mainly English. The participants received monetary compensation for their participation. Only the students who reported normal hearing abilities took part in the experiment.

1.3.2 Materials and Design

All of the target words were chosen from Aralex, the lexical database for Modern Standard Arabic, (Boudelaa & Marslen-Wilson, 2010). Aralex is a database for MSA and does not provide information about any CA dialect. It contains very few flagged CA words. Aralex is built on written texts, primarily newspaper texts. The database provides information about the surface frequency as well as the type and token frequencies of roots and stems for 80,330 Arabic words based on a modern text corpus of 40 million words. One hundred sixty target words were selected to be used in the experiment. Half of the target words were highly frequent words while the other half were chosen from the low-frequency range. To the best of my ability, I manipulated the high-frequency and the low-frequency ranges of words to include subsets of words with large and small family sizes as well as with high and low cumulative frequencies for both their roots and stems.[8] The words were also varied in the number of morphological competitors (i.e., the number of morphological members that have higher frequencies than the target word) for both their roots and stems. Table 1.1 shows the ranges for the main seven continuous variables: word frequency, cumulative root frequency, root family size, cumulative stem frequency, stem family size, and the morphological cohort competitors for both types of morphology. The token frequencies of words (i.e. their surface, root or stem frequencies)

---

[8] Aralex provides frequencies for two types of stems: frequencies for unpointed stems, where short vowels and other diacritics are not written (e.g., سلم 'ladder', 'peace', 'to submit', 'to escape danger'), and frequencies for pointed stems, where diacritics appear above or below the word (e.g., سُلَّم 'ladder' سِلْم 'peace' سَلَّم 'to submit' سِلْم 'to escape danger). Aralex finders faced the challenge of the absence of diacritics from the corpus by developing a novel automatic technique with a set of concatenation and disambiguation rules. This technique helped the database finders choose the correct vowelled/pointed solution for each ambiguous (i.e., unpointed) orthographic form in the corpus among the other several ponied-alternatives listed in AraMorph (Buckwalter, 2002). Since the experiment presented in this chapter is an auditory rather than a visual experiment, the frequency of the phonological (i.e., pointed) stem becomes more relevant.

are computed as the rate of their occurrence counts per one million words. For example, a surface frequency of 0.03 means that the given word has occurred only once in the 40-million-word corpus, a surface frequency of 0.05 means that the given word has occurred twice in the 40-million-word text, and so forth.

Table 1.1: *Descriptive statistics for the tested variables*

Variable	Min.	Max.	Mean	S.D.
Surface word frequency	0.03	153.90	11.40	25.4
Cumulative root frequency	3.32	8294	988.40	1454.51
Root family size	2	53	18.33	11.10
Root morphological competitors	0	20	3.21	4.28
Cumulative stem frequency	0.10	1512	149.50	278.91
Stem family size	1	8	3.40	2.12
Stem morphological competitors	0	6	0.57	0.93
Uniqueness point	233	977	551.4	130.96
Word duration	437	1188	754	126.5

The output frequencies of the database were not always identical to the final frequencies used in this experiment as the database frequencies underwent another stage of screening. In general, there are four reasons for the possible discrepancies between the database values and the actual values used in this study. First, sometimes the database lists loan words and foreign names under Arabic root entries. These loan words and foreign names were excluded manually. Second, the database calculates stem frequencies primarily based on the inflectional morphology of the word. This count, however, does not match with the more comprehensive definition of the cumulative stem frequency used here, which covers frequencies of all words that share the same stem. As a result, I added together the inflectional frequencies for all possible derived words. For example, the stem frequency of the participle *mu-ʃaarak* ‘being participated (in)’ was manually added to the stem frequency for the perfective verb *ʃarak-a* ‘participated’. Similarly, the stem frequency of *dʒawhar* ‘essence’ was added to the stem frequency of *dʒawhar-y* ‘essential’. Third, sometimes the stem frequencies do not appear directly under the word entry. In such cases, I looked under the root entry of the word. Fourth, some Arabic words vary in their spelling or are frequently misspelled. The same Arabic word [ʔadʒdʒadʒa] *‘to inflame’* appears in two separate lexical entries, one with the letter that corresponds to the glottal stop (أجج) and one without it (اجج). Since the present study is about the spoken modality of word recognition, I added frequencies of both possible spellings together to be the surface frequency of the word.

One hundred fifty-three words of the selected items had tri-consonantal root and seven words had quadri-consonantal roots. Based on the stem-based view of Arabic morphology, the tested words consisted of 83 derivationally complex words, both prefixed and suffixed, and 77 simple (i.e., monomorphemic) words. The target words belonged to three word classes: 87 nouns, 37 verbs, and 32 adjectives, whereas 4 words were ambiguous between nouns and adjectives. I also measured the selected words in terms of their uniqueness points (UP) and durations in milliseconds, and included them as nuisance variables in all statistical analyses. UP is the point of the word at which there is only one possible candidate left in the cohort of the other competing words (Marslen-Wilson, 1987). For more information about how UPs of the target words were measured, see Section 4.4.2 of this thesis. Word duration is the acoustic time distance between the onset and the offset of the spoken item. Both traditional UP and word duration have proved valid for predicting spoken word recognition in a considerable body of research (e.g., Balling & Baayen, 2008; 2012; Wurm, 1997). The experiment also included 160 pseudowords (i.e., nonsense words). The pseudowords were constructed by changing one or two phonemes of real MSA words. The pseudowords were phonologically and phonotactically legal in MSA. They were also comparable with the real word in the number of phonemes, syllables, and duration.

I transformed surface, root, and stem frequencies to their logarithmic scales to normalize the skewed distributions and minimize the effect of atypical outliers. Table 1.2 summarizes values of correlations between the tested lexical variables. This shows that collinearity between some of the lexical variables is still high despite their careful selection.[9] It indicates that log stem frequency is highly correlated with log surface word frequency. Log root frequency is strongly correlated with root family size and root morphological competitors. Both morphological families are correlated with their morphological competitors. UP is also highly correlated with word duration.

The problem of collinearity between continuous independent variables was unavoidable in most of the word recognition studies. With high collinearity it becomes

---

[9] Some researchers use 0.5 as the lower limit for strong correlations (Kuperman, Bertram, & Baayen, 2008). Others consider a correlation strong when it exceeds 0.3 (Bürki & Gaskell, 2012). In this study, I adopted an intermediate position and identified the limit for strong correlations by 0.4.

difficult to assess which of the tested variables have actual effects on the dependent variable and which do not. Some researchers (e.g., Balling, 2008; Balling & Baayen, 2008) suggest residualizing and principle component analysis (PCA) as useful tools to reduce collinearity.[10] In a more recent work, Wurm and Fisicaro (2014) argue against these techniques, demonstrating that such procedures do nothing to solve the problem of collinearity and do not even change the main results. What residualizing and PCA actually give is at least one predictor that is almost impossible to interpret. Although the problem of collinearity is an unalterable fact of life, it is still less serious in the current study and can be ignored. This is because the entire point of this experiment is to talk about the percentage of unique variance explained (i.e., whether the additional explained variance is worth increasing the complexity of a model) rather than to decide whether the factors are significant or not based on their *p*-values. Some researchers (e.g., Cohen, 2012; Jaeger, 2010) employed a similar technique, testing whether a complex model is statistically justified compared to simple models, and noticed that such a method is "robust against collinearity" (p.37), and renders collinearity less harmful (Friedman & Wall, 2005; Tabachnick & Fidell, 2007). Based on this, having each of the two compared factors in a different model and comparing their power to a third model with both factors makes us less uncertain about whether the effect is real.

---

[10] Residualizing is to regress one variable against another correlated variable, and to take the difference between their actual and fitted values as a new factor alternative to the old one.

Table 1.2: *Correlation matrix between the nine variables. Correlation values larger than 0.4 were considered high and printed in bold.*

Variable	Log Word Frequency	Log Root Frequency	Root Family	Root Competitors	Log Stem Frequency	Stem Family	Stem Competitors	UP	Duration
Log word frequency	-	0.21	-0.07	-0.37	**0.60**	-0.06	-0.30	-0.01	-0.04
Log root frequency	0.21	-	**0.51**	**0.45**	0.38	0.03	0.01	0.08	-0.02
Root family	-0.07	**0.51**	-	**0.61**	-0.06	0.01	-0.03	0.26	0.02
Root competitors	-0.37	**0.45**	**0.61**	-	-0.32	0.00	0.13	0.14	0.08
Log stem frequency	**0.60**	0.38	-0.06	-0.32	-	0.21	0.18	-0.06	0.01
Stem family	-0.06	0.03	0.01	0.00	0.21	-	**0.48**	0.12	0.13
Stem competitors	-0.30	0.01	-0.03	0.13	0.18	**0.48**	-	0.05	-0.01
UP	-0.01	0.08	0.26	0.14	-0.06	0.12	0.05	-	**0.47**
Duration	-0.04	-0.02	0.02	0.08	0.01	0.13	-0.01	**0.47**	-

1.3.3 Procedure and Apparatus

The stimuli were recorded in a sound-attenuated booth with a PMD660 Marantz digital voice-recorder. The speaker was a male native speaker of Jordanian Arabic (not the author), who has received his school education in the Arabic language. The participants' response times (RTs) were measured in milliseconds, from the onset of each target word, with a SONY laptop PC (CPU 2.40 GHZ) running Windows 7 and E-prime 2.0 presentation software (Psychological Software Tools, Inc, Pittsburgh, PA, USA; http://www.pstnet.com). E-prime randomized every session block so that none of the participants had exactly the same order of words. The participants listened to the stimuli through Sony MDR-NC8 headphones.

The participants were asked to perform a single lexical-decision task: to respond as quickly and accurately as possible by pressing the YES key if the stimulus was a real word and the NO key if it was a nonword. The participants were instructed to use the index finger of their dominant hand to press the YES button and the index finger of the other hand to press the NO button. A participant could not listen to the next stimulus until he/she pressed one of the two buttons. After every response there was a 2000 ms interval of silence. Participants were able to take a break anytime by pressing the *space* key. However, they did not receive any RT or accuracy feedback during the experiment. The experiment started with 10 practice trials followed by the test items. Each experimental session lasted approximately 20 minutes. The participants were tested individually in a quiet room.

1.4 Results

One participant was excluded from the data analysis because of his/her high error rate (errors > 30%). Two words were also removed from further analysis because approximately half of the participants incorrectly judged them as nonwords. One of them was mispronounced, whereas the second word was rejected as a real word for an unknown reason. When RTs were plotted across participants, I found 1500 and 650 ms good upper and lower limits between responses and obvious outliers. Thus, RTs longer than 1500 or shorter than 650 ms were excluded from the analysis. These removed outliers were 2% of the total remaining data.

To answer the research questions and test the postulated hypotheses, I submitted the data to multilevel regression analyses. I created a sequence of mixed-effects models in R, version 3.2.3 with the *lme4* (Bates, Maechler, Bolker, & Walker, 2016) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2016) packages. The fitted multilevel-regression models included reaction times, measured in milliseconds, as the dependent variable. They also had logged word frequency, logged root frequency, logged stem frequency, root family size, stem family size, root morphological competitors and stem morphological competitors as the main independent variables. To remove possible background noise, the regression models included word durations and uniqueness points as nuisance variables. The fitted models incorporated subjects and words as two random variables. I used the *Akaike Information Criterion* (AIC; Akaike, 1974) and the *Bayesian Information Criterion* (BIC; Schwarz, 1978) as measures of how well a model fits its data. The smaller their values are, the better the fitted model will be. Both AIC and BIC are calculated almost identically. The only difference between them is how much they penalize a model for its complexity (Wagenmakers, 2007). BIC probably penalizes complexity a little too much; AIC probably does not penalize complexity enough.

I fitted an initial null model with the random variables only, and then I started to add the relevant variables and eliminate the irrelevant ones, using forward-stepwise and backward-stepwise comparisons. This was the procedure followed throughout the analysis until I found the most complex model that could account for the largest amount of variance in the data. The step-forward/backward method tells us what variables to include and what variables not to include before we move to test a higher complex model in the hierarchy. For example, I fitted a simple model with cumulative root frequency, another simple model with cumulative stem frequency, and a third nested model with both variables. Then I selected the model with the highest explanatory power (i.e., with the lowest AIC value) and included the other variable in the same way (i.e., one at a time). After every step of variable addition, I checked whether the latest variable(s) I added was/were not rendering the early variables irrelevant (i.e., did not improve the AIC/BIC values).

Figures 1.2-1.4 show that the correlations between reaction times and the three measures of token frequencies are highly noisy. Nonetheless, the correlations are still significantly linear. If we go from the average of surface word frequency to one standard deviation above that average, RTs will drop by 63 ms, as Figure 1.2 depicts. Similarly, one positive standard deviation change in the scale of cumulative root frequency speeds up RTs by 54 ms, as Figure 1.3 illustrates. The analysis also revealed that one positive standard deviation change in the magnitude of cumulative stem frequency corresponds to 48 ms average decrease in participants’ RTs. The comparisons show that the effect size for the stem frequency, cumulative root frequency, and cumulative stem frequency are very close to each other. Moreover, it seems that there is a possible ceiling effect for the cumulative root frequency. Figure 1.3 shows that the participants did not improve their RTs any further beyond the frequency of seven on the logarithmic scale of cumulative root frequency.

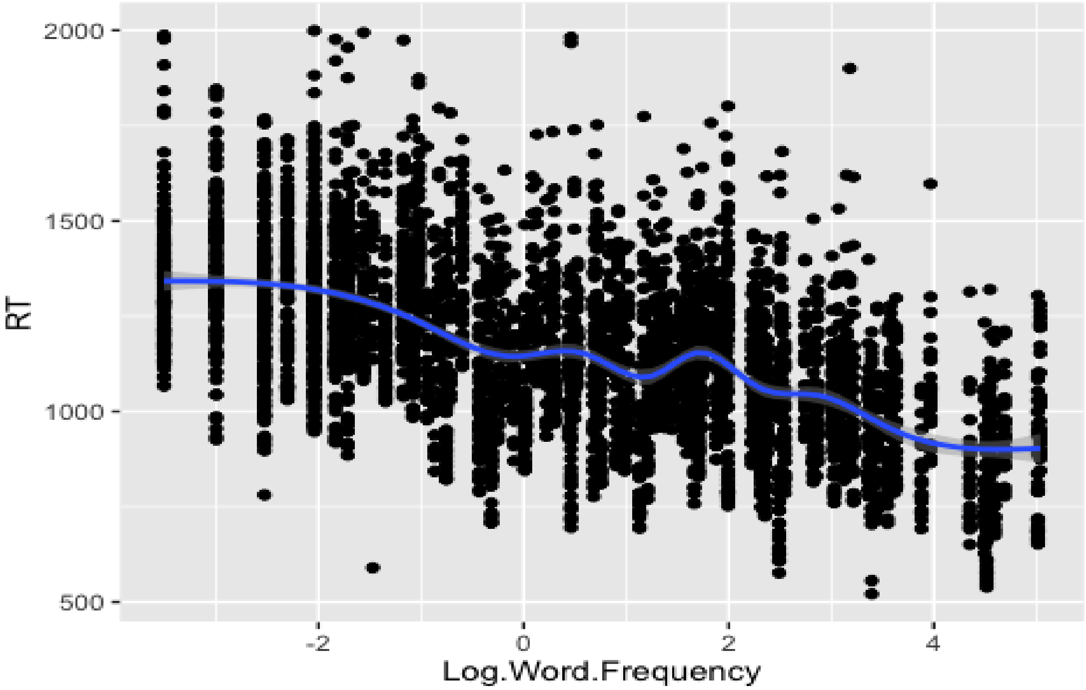


Figure 1.2: *Effect of surface word frequency on RTs.*

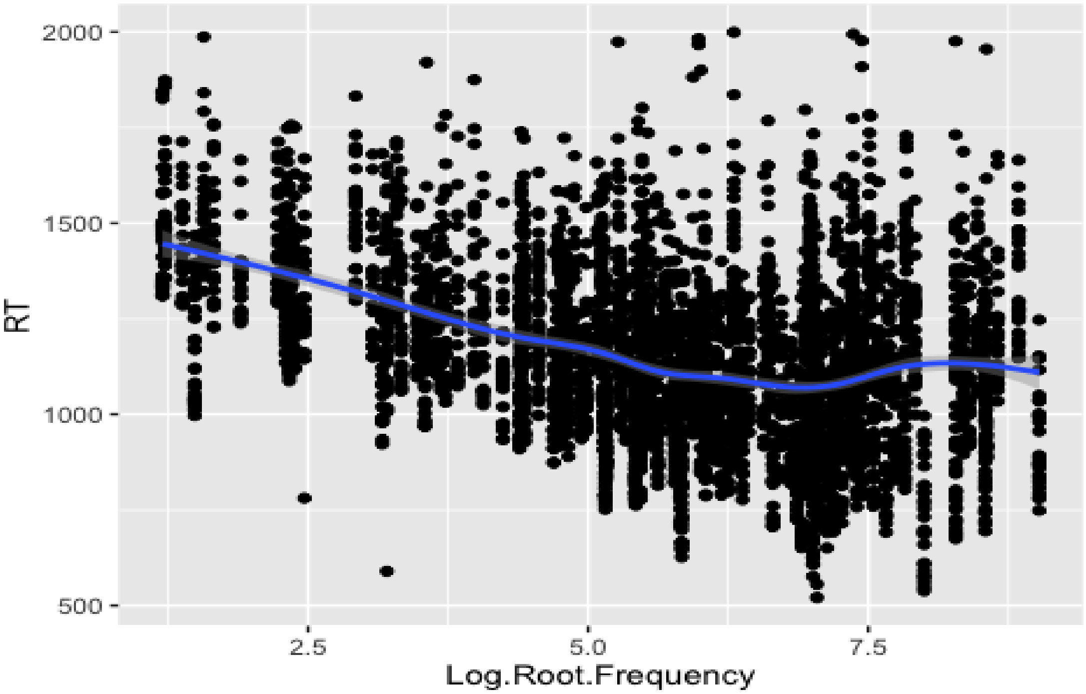


Figure 1.3: *Effect of cumulative root frequency on RTs.*

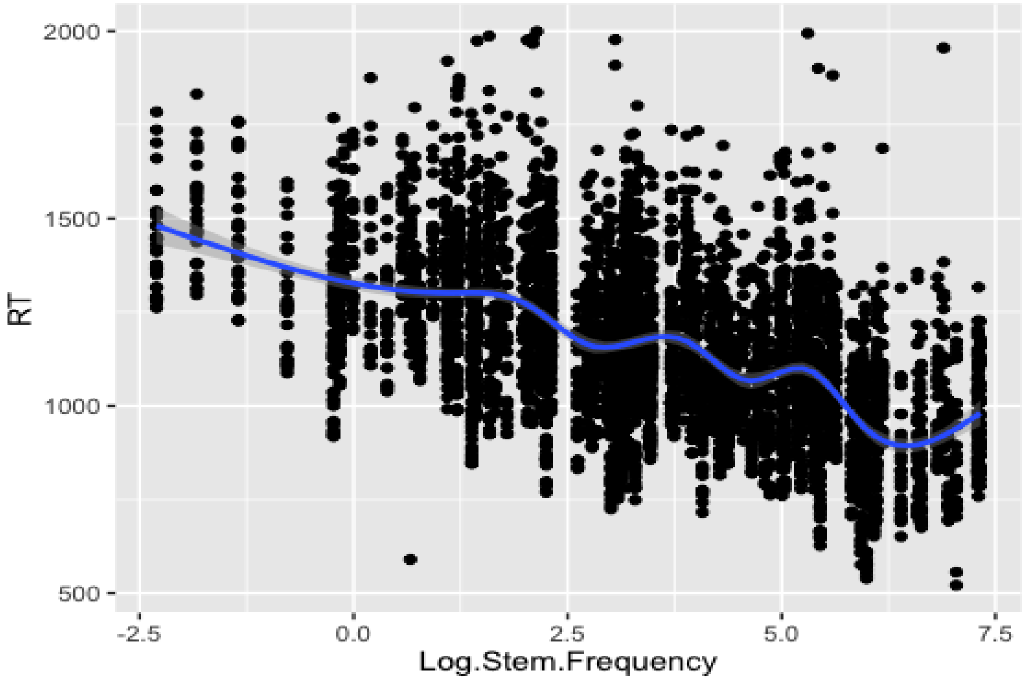


Figure 1.4: *Effect of cumulative stem frequency on RTs.*

Inferential statistics, summarized in Table 1.3, revealed that word frequency, root frequency, stem frequency, root-family size, and root morphological competitors have significant effects on RTs and the model that embraces all of these variables can account for more data variance. To justify why all of these variables are necessary for predicting RTs, Table 1.3 shows that leaving out any of them results in higher AIC/BIC values. This means that excluding any of these variables renders the model less predictive and powerful.

Table 1.3: *Comparisons between the final model (highlighted) and the simpler models***,** with the effect signal of the excluded variable.

Simple models	AIC	BIC	Effect Signal
UP/word duration/word freq/root freq/stem freq/root size+ root competitors	57321	57391	
UP/word duration/word freq/root freq/stem freq/root size	57406	57470	+
UP/word duration/word freq/root freq/stem freq/ root competitors	57399	57463	-
UP/word duration/word freq/stem freq/root size/root competitors	57478	57542	-
UP/word duration/word freq/root freq/root size/root competitors	57461	57525	-
UP/word duration/ root freq/stem freq/root size/root competitors	57561	57625	-
UP/word freq/root freq/stem freq/root size/root competitors	57527	57591	+
Word duration/word freq/root freq/stem freq/root size/ root competitors	57463	57527	+

These results gave positive answers to research questions one through five, asking about the effect of word frequency, root frequency, root family size, root morphological competitors and stem frequency on RTs. They were also consistent with the first three hypotheses, though the third hypothesis was partially validated because the stem effect was confined to the cumulative frequency. Moreover, a comparison of the final model against models with the other two remaining variables (i.e., stem family size and stem morphological competitors) revealed that the nested models were penalized for their extra complexity (acquired higher AIC/BIC values). Thus, adding any of these variables to the model did not improve its explanatory power, as Table 1.4 shows. These findings provided research questions six and seven, asking about any explanatory effect for stem family size and stem morphological competitor, with negative answers.

Table 1.4: *Comparisons between the final model (highlighted) and the more complex*

Complex models	AIC	BIC	Effect Signal
UP/word duration/word freq/root freq/stem freq/root size/ root competitors	57321	57391	
UP/word duration/word freq/root freq/stem freq/root size/ root competitors / **stem size**	57321	57398	**-**
UP/word duration/word freq/root freq/stem freq/root size/root competitors /**stem competitors**	57322	57399	+

models, with the effect signal of the added variable *(in bold).*

Despite the higher explanatory power for word frequency and cumulative root frequency, relative to cumulative stem frequency (compare their AIC/BIC values in Table 1.3), the three frequency measures had approximately equal effects, as figures 1.2-1.4 suggest. This gives a neutral answer to research question eight, inquiring about which frequency has a stronger effect on RTs. Based on this analysis, it is safe to say that there is not enough evidence to support either the fourth or the fifth hypothesis in the present study.

Table 1.5: *Effect of stem frequency in Arabic polymorphemic words.*

Model with stem freq.	AIC	BIC	Model without stem freq.	AIC	BIC
UP/word duration/word freq/root freq/ root size/ root competitors/ stem freq.	27662	27724	UP/word duration/word freq/root freq/root size/ root competitors	27723	27780

Table 1.6: *Effect of stem frequency in Arabic monomorphemic words.*

Model with stem freq.	AIC	BIC	Model without stem freq.	AIC	BIC
UP/word duration/word freq/root freq/root size/ root competitors/stem freq	29671	29734	UP/word duration/word freq/root freq/ root size/ root competitors	29748	29805

Based on the stem-based perspective of Arabic morphology, the tested words varied in their morphological complexity (i.e., polymorphemic vs. monomorphemic) and affix types (i.e., suffixed vs. prefixed). As a result, it was useful to examine whether both parameters of morphological complexity were sensitive to the effects of cumulative stem frequency. Tables 1.5-1.6 demonstrate that stem frequencies of both polymorphemic and monomorphemic words contributed to the explanatory power of their models (i.e., had lower AIC/BIC values), relative to models without them.

1.5 Discussion

One motivation for the present work was to assess the role of morphology in the lexical representation and processing of Arabic. Previous research used the priming techniques to examine the effects of Arabic morphology on lexical processing. However, this study examined the relevance of Arabic morphology through a simple-lexical decision task, where the target words were not influenced by any morphologically-related primes. Another motivation for this research was to evaluate the effects of two possible Arabic morphemes directly; namely, consonantal roots and stems (i.e., strings of consonants and vowels). The extant research presents the psychological effects of discontinuous morphology as evidence against continuous morphology (Boudelaa, 2014, Boudelaa & Marslen-Wilson, 2015) and the absence of such effects as evidence for continuous morphology or against any morphological role (Abu-Rabia & Awwad, 2004). The present work addressed the effects of both types of morphology directly in an attempt to give a clearer picture of their psychological status in the minds of Arabic speakers.

This study raised and tried to answer three main questions. The first question addressed the relevance of Arabic morphology in processing the spoken words of Arabic.

It asked whether native speakers of Arabic process a spoken Arabic word as a single lexical unit, as sublexical morphemic units, or both. The second question asked whether Arabic lexical representation consists of roots only, stems only, or whether both types of morphemes are cognitively valid units in the mental lexicon of Arabic speakers. The third question was contingent on the answer to the second question. That is, if it turned out that Arabic morphology is both root-based and stem-based, which of them has a more robust status in the minds of Arabic lexical processors? To answer these questions, I selected a sample of MSA words varied in their whole-word frequency, root-based frequency, and stem-based frequency. I also manipulated different morphological competitor and family sizes for both proposed root-based and stem-based morphemes to give more or less decisive answers to the research questions.

The results showed that each of the three token-frequencies contributed to word processing. The statistical model that comprised the whole-word frequency, the cumulative root-based frequency, and the cumulative stem-based frequency accounted for more data than the statistical model(s) that excluded any of them. The research results revealed that the three frequency counts had nearly equal positive effects on word processing. Shorter response latencies related to words with more common surface constituents (by 63 ms in one standard deviation change), words with more common root constituents (by 54 ms in one standard deviation change), and words with more common stem constituents (by 48 ms in one standard deviation change). The three token-frequency effects are the crucial findings for the claim that Arabic words, stems, and roots are approximately equally important in word processing. These findings agree with the Indo-European studies that reported facilitative effects for stem token frequencies in both visual modality (e.g., Burani & Thornton, 2003; Schreuder et al., 2003) and auditory modality (e.g., Meunier & Segui, 1999). The results also concur with the facilitative effects of surface word frequency reported in many works on word recognition (e.g., Baayen et al., 2007; Cleland et al., 2006; Meunier & Segui, 1999; Wurm, 1997; Wurm et al., 2006).

The findings revealed a positive facilitative effect for root family size and a negative inhibitory effect for their morphological cohort competitors. Both effects were less robust than the effect of root frequency. The larger the family size for the Arabic word, the shorter the RTs would be. Previous works reported similar facilitative effects for root morphological size in Hebrew (Moscoso del Prado Martin et al., 2005) and Arabic (Boudelaa & Marslen-Wilson, 2011). Nevertheless, words with less frequent root neighbors were reacted to faster than words with more frequent root neighbors. The inhibitory effect of root competitors concurs with the inhibitory effect of French stem competitors reported in Meunier and Segui’s work (1999). In this work, the family size effect was found with root morphological neighbors but not with stem morphological neighbors, as the morphological size of the stem did not strengthen the explanatory power of the statistical model. Given this morphological gap, the role of root-based morphology is clearer than the role of stem-based morphology.

This study adds to the growing body of research showing that Arabic root constituents are valid predictors for lexical decisions. The results are in accordance with the findings of the priming experiments that revealed psycholinguistic validity for Arabic roots, independent of form or semantic transparency (Boudelaa & Marslen-Wilson, 2004 b; 2007; 2015; Mahfoudhi, 2005). It is still possible to attribute the stem frequency effect found in the experiment to the phonological (i.e., form) similarities, per se, rather than to the morphological relatedness between words. Yet, I can argue against this assumption based on some other findings. Previous research reported that form overlap between words produces negative effects on the lexical decision (i.e., longer RTs) when words share a semantically unrelated string of sounds (e.g., Drews & Zwitserlood, 1995; Feldman & Andjelcovic, 1992). For example, Drews and Zwitserlood (1995) found an inhibitory effect on German target words *(e.g., keller* 'basement'*)* when presented with orthographically-related primes *(e.g., kelle* 'ladle'*)*. Accordingly, if the effects of stem frequencies found in the present work were phonological, there would have been inhibitory effects when a target word activated its assumed phonological neighbors. As opposed to this, the current study revealed a positive effect (shorter RTs) on cumulative stem frequencies. Moreover, Baayen et al. (2006) found that the role of form frequency is very marginal, compared to the role of morphological and surface frequencies, in both naming and visual lexical-decision tasks. However, results of the present experiment revealed a reasonable explanatory power and a comparable effect size for cumulative stem frequencies relative to root and surface frequencies. Coupled with Drews and Zwitserlood, and Bayeen's findings, the present results prefer the morphological account of Arabic stems over the possible phonological account. The partial stem effect found in the present work agrees with the proposal that Arabic has concatenative morphology (Abu-Rabia & Awwad, 2004; Benmamoun, 1999, 2003; Ratcliffe, 1998, 2004), but without dispensing with the nonconcatenative morphology of Arabic (Boudelaa & Marslen-Wilson, 2015).

This work revealed two other findings related to the stem-based account of Arabic. First, the cumulative frequencies of the proposed Arabic stems were attested to both suffixed words (e.g., *ʔusluub-yyah* ‘stylistic’) and prefixed words (e.g., *ta-badara* ‘to cross someone’s mind) when other frequencies were held constant. The finding agrees with the prefix-stripping hypothesis in visual word recognition (e.g., Giraudo & Grainger, 2003; Holmes & ORegan, 1992; Taft, 1979) and its spoken modality (e.g., Wurm, 1997). The results, on the other hand, did not support the view that cumulative stem frequency is irrelevant in prefixed word processing because these words do not undertake prefix-stem parsing and are accessed as a whole unit (e.g., Colé et al., 1989; Meunier & Segui, 2003). Second, the stem frequency effect proved valid for both monomorphemic (e.g., *naaʃada* ‘to entreat’) and polymorphemic (e.g., *xaʃabiyy* ‘wooden’) Arabic words. This finding is in line with the proposal that the cumulative stem frequency is relevant in both morphologically simple words (e.g., Baayen e al., 2006; Ford et al., 2003; Schreuder & Baayen, 1997) and morphologically complex words (e.g., Baayen et al., 2007; Burani & Thornton, 2003; Meunier & Segui, 1999). It also suggests that morphologically simple and complex words could undergo a very similar process of recognition (Moscoso del Prado Martin et al., 2004).

In the following brief discussion, I will suggest some possible accounts of why literate speakers of Jordanian Arabic have two types of morphology in their lexicon. The first hypothesis is that both stems and roots co-exist as constant and static entities of Arabic morphology independent of any external factors. According to this assumption, native speakers of Arabic naturally acquire and develop both types of morphology. The second possible reason is that Arabic, like English and most other human languages, is a stem-based language, and the root/pattern activation is just the result of literacy. The Arabic language curricula of different stages focus on the idea that Arabic words are built of consonantal roots and word patterns. School instructions may parallel the assumed discontinuous morphology to the universal mental concatenative representation of morphology. However, Boudelaa and Marslen-Wilson (2013) found that both MSA and the Tunisian dialect of Arabic have similar cognitive properties: both show an equal priming effect for their roots and an equal priming effect for their word patterns. The authors introduced this finding as evidence for the autonomy of the root/word pattern from any literacy effect since regional dialects of Arabic are not used in literacy. Here, we can easily assume that literacy might have reconstructed both varieties of Arabic even though Tunisian Arabic is not the language of reading and writing. That is, literacy reshapes the linear morphology of MSA, which also can be generalized to reconstruct the morphology of CA, by virtue of their shared roots and/or word patterns, and Arabic speakers' awareness of the analogy between the two varieties of Arabic.

A third account is to assume that root-based morphology is the default unmarked Arabic morphology, whereas stem-based morphology emerged from participants' good proficiency of English. In Jordanian public schools teaching English as a foreign language occurs in grade 1 through grade 12. In addition, English is the medium of instruction in many faculties and departments of Jordanian universities. As a result, student awareness of stem-based morphology of English may carry consequences for lexical representation and processing of Arabic. There is some evidence for restructuring transfer (i.e., the incorporation of second language (L2) elements into first language (LI) that results in some L1 structural changes) across many languages (e.g., Cook, 2003;

Pavlenko, 2000; 2003; 2004), and Arabic is a case in point (Qasem & Foote, 2010). Forster and Deutsch's (1997) claim that priming effects are motivated by morphological relatedness rather than phonological similarities between the prime and the target in nonconcatenative languages, including Arabic. Qasem and Foote (2010) tested Forster and Deutsch's hypothesis on two groups of native speakers of Arabic with different levels of L2 English proficiency. The authors found increasing phonological priming effects (e.g., *shoulder-kahif* 'cave') with higher L2 English proficiency. Qasem and Foote demonstrated that with the increasing English proficiency, the participants activated the Arabic phonologically-related words *kahif* 'cave' along with the Arabic translation of shoulder, *katif*.

To decide between these three possible accounts of dual morphologies, we may need to know how illiterate monolinguals of Arabic, early school dropouts, react to both types of morphology when they listen to their native CA words.[11] Such an investigation can lead to one of three possible scenarios. There might be effects for both morphological frequencies. This result would indicate that Arabic is inherently stem-based and root-based, and both types of morphology have a static status in the minds of native speakers of Arabic. The second scenario can be a stem-based frequency effect only. This proposes that lexical representation is universal (Berent, Vaknin, & Marcus, 2007), and that root-based morphology is motivated by literacy. The last scenario is to find a root-based effect only. In this case, it is legitimate to say that consonantal roots are the default morphological representation of Arabic words (e.g., Boudelaa & Marslen-Wilson, 2013),

---

[11] This investigation will be possible if a database, similar to Aralex, is built for any colloquial variety of Arabic. Text messages and scripts of informal TV shows, soup operas and movies can be the basic material for the suggested database.

and the stem-based effect found in the present study is an instance of reverse transfer from English into Arabic.

Results of this research can be used to evaluate six main models of word recognition. The present findings cannot support *the full-listing hypothesis* of lexical access (e.g., Sereno & Jongman, 1997), in which only full-word frequency can influence lexical access. Panel A of Figure 1.5 represents *the full-listing model*. The results are also inconsistent with the other opposite extreme model, the *full-decomposition model* depicted in Panel B of Figure 1.5. This model claims that sublexical information (e.g., letter, phonemes) is mapped onto meaning via morphemes and no complex word is stored or accessed. According to this model, only stem/root frequencies can influence reaction times to words.

As effects of both full-word and morpheme frequencies prevail in this study, we can consider the implications of the results in terms of some mixed models of lexical access. Panels (C-E) illustrate three possible mixed models of word recognition. I imposed a root level between the stem and the word representation in model architectures. The reason for this extra tier is to make it possible to account for both morphological roles in word processing. Before embarking on any evaluation, it is necessary to understand a major difference between spoken and visual word recognition. Speech has temporal components and speech signals unfold in time. Listeners determine what the word is based on the continuous mapping of sequential inputs onto word representation. This is known as the *cohort model* of spoken word recognition (Marslen-Wilson, 1987; Warren & Marslen-Wilson, 1987). In visual word recognition, on the other hand, information about the word can be available to the reader all at once. Therefore, an accurate description for any possible model in light of spoken data should take the *cohort model* into account.

While the current research is not designed to decide on which of these three representations is the most proper one, each model can account for part of the findings. For example, a sublexical model (e.g., the *direct access* or the *obligatory decomposition*), visualized in panel C of Figure 1.5, can account for the decomposition of the suffixed words, as stem inputs precede the whole word inputs in the spoken domain. Sublexical models cannot account for the decomposition of prefixed words because, according to the cohort theory of spoken word recognition, the stem of a prefixed word cannot be accessed before the whole-word representation. Alternatively, the decomposition of prefixed words is consistent with the supralexical model outlined in panel D. The supralexical model proposes that stems are accessed as soon as or directly after the whole words. This is compatible with the position of stems in prefixed words. For both suffixed and prefixed words, their roots are accessed after their stems, as roots are a kind of abstraction over the stems. These two models can be incorporated together to form a dual model similar to the *morphological race model* (e.g., Burani & Caramazza, 1987; Schreuder & Baayen, 1997). For this model, see panel E of Figure 1.5, stem frequency, root frequency, and word frequency are the result of the repeated access to the same stem, root and word. Which entity succeeds (i.e., is identified first) depends on when the unit inputs become available to the listener. While root inputs always become available after the stem access, the earlier access to stems or whole words depends on whether the word is prefixed or suffixed.

A distributed connectionist model similar to the one proposed by Plaut and Gonnerman (2000) can also account for the present findings. According to the distributed connectionist approach, words are not single nodes but the results of the co-activation of distributed phonological, orthographic, and semantic features. A connectionist network is trained to map the form and meaning consistency of the morphologically complex words through hidden units, which mediate between form and meaning (see panel F of Figure 1.5). This model is sensitive to form-meaning consistency (Gonnerman et al., 2007) and frequency (Davis et al., 2003; Seidenberg & McClelland, 1989). The network of words becomes stronger when their shared components are more frequent and more consistent. For example, the internal representation of a word such as *tasˤaaraʕa* 'fight each other' will not decompose into the stem [sˤaaraʕa] 'fight' and the prefix [ta-] 'reciprocal', or into the tri-consonantal root {sˤrʕ} and the word pattern {tafaaʕala} 'the reciprocal pattern'. Instead of that, there will be a consistent pattern of online activation over several nodes. That is, when *tasˤaaraʕa* is processed, the word *musˤaaraʕah* 'wrestling' will also be activated by virtue of their shared stem [sˤaaraʕa] (i.e., both words recruit [sˤaaraʕa]). Moreover, some units involved in the representation of the stem [sˤaaraʕa] will also be involved in the representation of the stem [sˤarʕ] 'epilepsy'. These units are the shared roots. The two words *tasˤaaraʕa* 'fight each other' and *sˤarʕah* 'crazy, rage' are linked through their tri-consonantal root {sˤrʕ}, which activates the other related word(s) while one word is being processed. Among other relevant factors, the strength of the network between these Arabic words depends on their stem and root frequencies.

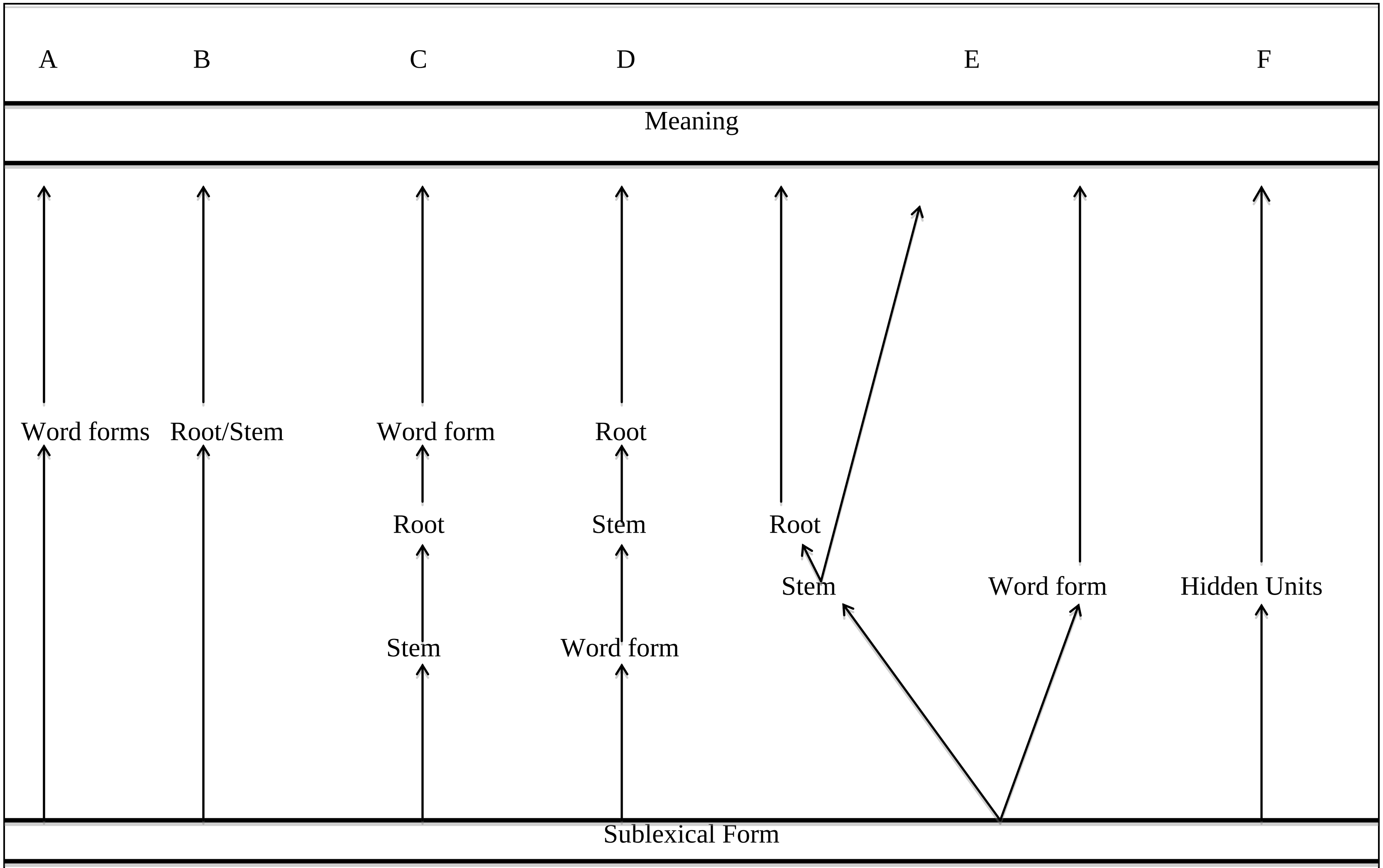


Figure1.5: *Six architectures of word recognition with and without morphology*. *Adapted from Diependaele et al. (2012)*

One limitation of the present work is that estimates of roots and stems were highly noisy, as figures 1.3-1.4 of the results section show. This noise could partially be attributed to some inherently unavoidable problems in the mechanism of the database. It is also possible that the frequency measures for the MSA words used in this experiment are partially influenced by JCA. That is, most of the target words have cognates and near cognates in JCA, and their stems and roots may have different frequencies in the other variety of Arabic.

In conclusion, the research reported here shows that Arabic lexical decisions are not the result of processing a single level of representation, but rather the result of activating multiple sources of information. These resources include, but are not limited to, information from the storage of stems, information from the storage of consonantal roots, and information from the storage of the whole words. The experiment showed that the effect of stem frequency is comparable to the effect of root frequency. In Arabic, roots recur more systematically and saliently than stems, which are usually less productive and organized. The advantage of Arabic stems, on the other hand, could emerge from their strong orthographic/phonological overlap relative to Arabic roots. It is probable that these unique characteristics of roots and stems equilibrate their effects on word processing. Alternatively, some researchers (e.g., Bat-El, 2003; Berent et al., 2007) propose one unified account of morphology. They indicate that what is beloved to be Semitic root-based morphology may reflect the quantitative difference in the vowel modification of stems, something that exists, but to a smaller extent, in many human languages. If this proposal happens to be correct, then putative roots are, in fact, stems with gradable vowel changes. This hypothesis is a subject for further research.

Chapter Two

Study II: Word Recognition in the Sentential Contexts of Arabic Diglossia: Is There a Switching Cost?

2.1 Introduction

One of the fundamental interests of bilingual research is to explore how bilinguals process their languages compared to monolinguals. It raises the question of whether and in what conditions bilinguals access one of their two languages with an advantage over the other. It also addresses the question of whether, and to what extent, bilinguals limit their lexical accessibility to the current relevant language without activating the vocabulary of the irrelevant language. Available research suggests that the answer to these questions depends on bilinguals' relative proficiency/experience in the second language and their expectation about what language they are going to perceive. Previous studies found that unbalanced bilinguals, who are unequally proficient or practiced in the two languages, usually process their first language (L1) more easily than their second language (L2; e.g., Kroll et al., 2002; Talamas et al., 1999; also see Chapter Three of this dissertation for more details). Other investigations found that bilinguals recognize target words in language-consistent contexts (i.e., an L1/L2 target word in an L1/L2 context) more easily than the same target words in language-inconsistent contexts (i.e., an L1 target word in an L2 context or an L2 target word in an L1 context, e.g., Groasjean, 1997; Proverbio et al., 2004; Soares & Groasjean, 1984).

The purpose of the study presented in this chapter is to determine whether the language-processing bias that concurs with classical bilingualism also occurs in the situation of Arabic diglossia, where speakers use two different varieties of the same language rather than prototypical independent languages. In the context of Arabic diglossia, Colloquial Arabic (CA) is the variety used in informal face-to-face conversations, whereas Modern Standard Arabic (MSA) is the typical variety of literacy and formal occasions (see section 2.2 of this chapter). Although each of the two varieties of Arabic has its domain of use, literate speakers of Arabic sometimes alternate between MSA and CA at the word, sentence, or discourse levels in the contexts of different levels of formality (Albirini, 2011; Bassiouney, 2006; Eid, 1988). Speakers of Arabic shift into MSA to construct high codes such as importance, high prestige, identity, seriousness, and sophistication. They also code switch into CA to construct low codes such as unimportance, low prestige, accessibility, and non-seriousness (Albirini, 2011).

One of the methods psychologists use to investigate bilinguals' bias in language-processing is to measure their reaction times (RTs) to target words of their L1 and L2 in both mixed-language and single-language contexts. For example, researchers can measure how fast English/French bilinguals respond to French words (e.g., cuillere) compared to English words (e.g., spoon). They also can measure how fast the same English/French group of bilinguals respond to switched words, words the lexicon of which is inconsistent with the lexicon of the preceding word, (e.g., in a mixed-language list such as: *deer, screen, <u>cuillere</u>, etoile, <u>man</u>, etc.*) compared to the same words in non-switched contexts (e.g., in a French list such as: *cerf, ecran, <u>cuillere</u>, etc.*; and in an English list such as: *dear, screen, spoon, star, <u>man</u>, etc.*). The idea is that if bilinguals respond to their L1 words faster than their L2 words, this suggests that the bilinguals under investigation are unbalanced bilinguals, having considerably less proficiency in L2 compared to L1 (Kroll & Stewart, 1994; Kroll et al., 2002). As far as the language switching is concerned, if bilinguals react to switched words in a mixed-language list more slowly than non-switched words of a single-language list, this suggests that bilinguals activate one of their two languages at a time or activate it more than the other language. Researchers usually refer to the extra time a bilingual may spend to respond to switched words compared to non-switched words as the switching cost. If there is a switching cost, this indicates that bilinguals almost turn off the lexicon of the irrelevant language, and retrieve it when they encounter a switched word, which is time-consuming. Alternatively, if bilinguals show no RTs difference when they process single-language compared to mixed-language stimuli, this indicates that bilinguals activate both of their languages equally all the time.

The study presented in this chapter investigated whether Standard Arabic is an L2 for native speakers of Arabic by measuring how fast they react to words of CA compared to words of MSA. It also examined the possibility of switching cost in Arabic diglossia by testing how fast native speakers of Arabic react to Arabic words in single-language conditions compared to the same words in code-switching conditions (i.e., alternating to a word from the other variety of Arabic within the same conversation). Moreover, the findings of this investigation will be used to evaluate the predictions of two different models of bilingual word recognition: the *bilingual interactive activation plus* (BIA+; Dijkstra &Van Heuven, 2002) and the *bilingual model of lexical access* (BIMOLA; Grosjean, 1997, 2008).

2.2 Background and Research Hypotheses

Native speakers of Arabic acquire their language in the special linguistic context of *diglossia* (Ferguson, 1959), where different spoken dialects of Arabic coexist with the more formal variety known as Modern Standard Arabic (Abu-Rabia, 2000; Maamouri, 1998). The two varieties show some divergences and convergences at all levels of linguistic description. Modern Standard Arabic (MSA) is used in formal situations such as public occasions, religious contexts, media, and press. It is no one's native language, and Arabs usually learn it in school.[12] MSA is the language of Arabic literacy (i.e., the language of reading and writing). Alternatively, the different varieties of colloquial Arabic (CA) are acquired early and used in everyday situations when people communicate with each other. Reading and writing in CA varieties is not common and limited to some aspects of social media (e.g., text messages, Twitter, Facebook; Al-Khatib & Sabbah 2008; Mostari 2009). In such contexts, native speakers of Arabic transliterate spoken colloquial Arabic words with MSA letters, Romanized letters or a combination of the two. While all native speakers of Arabic use the same MSA variety, their spoken dialects vary from one state to another and even in different regions of the same state. The diglossic situation of Arabic introduces CA as the dominant language of listening and speaking (i.e., face-to-face contact) and MSA as the dominant language of reading and writing.

The linguistic situation in Jordanian society is a typical example of Arabic diglossia (Al-Saidat, 1999; Suleiman, 1985). Despite the similarities between Jordanian

---

[12] This does not mean that native speakers of Arabic are not exposed to MSA before the school age. Some native speakers of Arabic experience MSA, from a more receptive standpoint, before they start formal schooling by watching MSA TV cartoons, early reading, and religion learning. However, it is beyond the scope of this study to examine any possible relationship between how early MSA is exposed to and the diglossic lexical access and representation.

colloquial Arabic (JCA) and MSA, the two varieties differ considerably in their phonology, phonetics, morphology, and lexicon (Al-Sughayer, 1990; Holes, 2004, Laks & Berman, 2014). The vocabulary of both varieties of Arabic converges and diverges in three aspects. First, the two varieties share many cognates or near cognates (i.e., inter-lingual words with strong form and meaning overlap). For example, the JCA *ruz* 'rice' and the MSA *ʔaruz* have similar forms and an identical meaning. Second, in many other cases, the two varieties of Arabic associate the same meaning with different forms. JCA words such as *tilifon* 'telephone' and *birwaaz* 'frame' correspond to MSA *haatif* and *ʔitˤaar*, respectively. A third group of words exists only in one of the two varieties, but their tri-consonantal roots are part of both lexicons. For example, the MSA word *malaabis* 'clothes' shares its tri-consonantal root {lbs} with other words from both varieties such as JCA *labiis* 'a tidy man'.

At the phonological level, MSA and JCA have very similar phonemic inventories. Yet, they differ in the existence of the voiced uvular stop /q/, the voiced velar stop /g/, and the voiceless palatal affricate /tʃ/. MSA has /q/ while JCA has /g/. In Fallahi (rural) and Bedouin (nomadic) vernaculars of Jordanian Arabic, /tʃ/ replaces /k/ in many words, and the interdental emphatic fricative /ðˤ/ replaces the alveolar emphatic stop /dˤ/. Additionally, /d/, /t/ and /s/ are allophones of the voiced interdental fricative /ð/ in Madani (urban) Jordanian Arabic. Moreover, MSA has three short vowels with their long correspondences /a/, /i/ and /u/ while JCA has two additional vowels: the mid-front vowel /e/ and the mid-back vowel /o/. These two mid-vowels replace the MSA diphthongs /ai/ and /au/, respectively. As for the syllable structures, the consonant clusters of MSA can only occur in the final position of the syllable (i.e., in the coda: e.g., *qahr*) but not in its initial position (i.e., in the onset). Alternatively, JCA words allow consonant clusters in the syllable-initial position (e.g., *ħ<u>raam</u>*) whereas the sequence of two consonants is uncommon in their codas.

Taking into account the linguistic and experiential differences between CA and MSA, it is possible that literate native speakers of Arabic are de facto unbalanced bilinguals. Ibrahim and Aharon-Peretz (2005) conducted semantic priming experiments on Palestinian Arabic/Hebrew bilinguals who learn both MSA and Hebrew in Arabic schools of Israel. The authors concluded that MSA constitutes an L2 rather than just a formal register of Arabic for native speakers of Palestinian Arabic. They built their conclusion based on two findings. First, the participants' RTs to MSA and Hebrew were slower than their RTs to Palestinian Arabic. Second, the semantic priming effects from Palestinian Arabic to MSA and Hebrew were stronger than the semantic priming effects from MSA and Hebrew to Palestinian Arabic. The researchers related these findings to the strong/direct link between L1 words and their concept compared to the loose link between L2 words and their concepts, which costs more processing time (for further details see Section 3.2 of Chapter Three). On the basis of the different ages at which the two varieties of Arabic are acquired/learned and the different sociolinguistic contexts in which they are experienced, we postulated the following hypothesis:

H1.2: Literate speakers of JCA recognize JCA words faster than MSA words.

The rest of this section will discuss the findings of the bilingual studies on mixed-language processing, and will introduce the second and third research hypotheses. Although the domain of this paper is spoken word recognition, I will review research from both auditory and visual word recognition. This is because, even though each input modality can have its own consequences for processing, their abstract underlying mechanism is almost the same (Dijkstra, 2005). That is, different researchers argue for different models of bilingual word recognition because they disagree on how bilinguals process their languages in general, rather than because they investigate different modalities of language. Caramazza and Brones (1980) reported no RT difference between words visually presented in single-language and mixed-language conditions when they asked a group of Spanish/English bilinguals to decide whether the target word (e.g., *apple*, *manzana*) belonged to a particular category (e.g., fruit). However, Grainger and Beauvillain (1987) used a lexical decision task (i.e., whether the stimulus was a word or non-word) to determine whether bilinguals process mixed-language stimuli with extra efforts. The authors found that French/English bilinguals judged visual target words in a single-language list more easily than the same targets when they follow words from the other language. However, the switching cost disappeared when the authors used word stimuli with orthographic cues unique to the target language. Grainger and Beauvillain interpret their findings as evidence for language selectivity: bilinguals deactivate the lexicon of the other language during word recognition. This deactivation transfers to the switched word, but bilinguals can quickly readjust the deactivation when the switched word contains language-specific cues.

Follow-up research on both language comprehension (Thomas & Allport, 2000; von Studnitz & Green, 2002) and production (Costa & Santesteban, 2004; Green, 1998; Macizon et al., 2012; Meuter & Allport, 1999) argues that switching cost is external, sensitive to the task demand and the sequential pattern of responses, rather than internal, sensitive to shifts in lexical activation. Thomas and Allport (2000) found that language switching interacts with the response of the previous trial in cue-induced language selection (i.e., decide whether the stimulus typed in red is a French word or not and whether the one typed in green is an English word or not). In repeated responses, the bilinguals reacted faster on non-switch trials and slower on switch trials. There was no significant difference between the size of the switch cost for words presented in the correct language context (i.e., Yes response) and words presented in the wrong language context (i.e., No response). Similarly, von Studnitz and Green (2002) found that their German/English participants, engaged in a semantic categorization task, reacted slower to language switched words only when they had to make two repeated responses (two consecutive animate or inanimate responses). All of this suggests that nothing unique in the linguistic systems of bilinguals is responsible for language switching costs. As any other non-linguistic stimuli, the switching cost of linguistic stimuli appears and disappears in compliance with the task demands and the persistence of the previous response (Becker, 2013; Hunt & Klein, 2002; Monsell, 2003).

As the previous discussion demonstrated, most of the findings on how bilinguals process their single languages relative to their mixed language come from investigating words in isolation. However, people rarely listen to or read single words. During their daily-life interactions, bilinguals listen to meaningful sentences where the base language usually makes up from 80% to 90% of the utterance (Grosjean, 1997). Thus, it is possible that the simultaneously balanced activation of both languages, represented in the absence of lexical switching costs, is because of the decontextualized nature of the stimuli. There are relatively few studies, with less decisive results, on how bilinguals process mixed language compared to a single language in sentential contexts (for some review see

Dijkstra, 2005; Lagrou et al., 2013). These studies will be the main focus of the following discussion.

Some recent research has proposed that the language of the carrier sentence does not provide bilinguals with enough cues to restrict their lexical access to the language of that sentence (e.g., Gullifer et al., 2013; Ibáñez et al., 2010; Schreier, 1998). For example, Schreier (1998) found no evidence for a processing delay when a group of German/English bilinguals reacted to German/English code switches in spoken sentence contexts. Gullifer et al. (2013) replicated this finding in visual word recognition of Spanish/English bilinguals. These results agree with the BIA+ model of word recognition (Dijkstra &Van Heuven, 2002). The BIA+ proposes that bilinguals always process their languages non-selectively, as they organize the vocabulary of the two languages in one unified lexicon. That is, candidate words from both languages compete with the target word for selection, regardless of the base language of the context. Accordingly, there is no top-down inhibitory effect from the prior language (i.e., a higher level) on any possible guest word (i.e., a lower level). However, other studies have argued for the effect of the base language on guest-word recognition (e.g., Grosjean, 1997, 2008; Proverbio et al., 2004; Soares & Grosjean, 1984). In a lexical decision task, Soares and Grosjean (1984) found that fluent Portuguese/English bilinguals responded faster to spoken Portuguese and English words embedded in a consistent language context compared to the same target words embedded in a different language context. In a more recent experiment, Domenighetti and Caldognetto (1999), as cited in Grosjean (2008), found that French/Italian bilinguals reacted with 46 ms delay when they named (i.e., listened to and repeated) Italian code switches in a semantically neutral French sentence relative to comparable French target words.

Grosjean (1997; 2008) suggested that recognizing guest words (i.e., words from the other lexicon) embedded in a base-language context (i.e., the dominant language of conversation) is a dynamic process governed by many involved factors. The first factor is related to the language mode of listeners. It refers to bilinguals' expectations about what they are going to hear: "The higher the expectation, the easier will be guest word recognition" (Grosjean, 1997; p. 241). At one endpoint, a bilingual can be in a monolingual language mode (i.e., listening to a monolingual speaker), where he/she activates only one of his/her languages. At the other endpoint, the bilingual can be in a bilingual language mode (i.e., listening to another bilingual), where he/she activates both languages. (For a critique of the concept of language mode, see Dijkstra et al., 2003). The listener can also be in a semi-monolingual mode, as language mode is a continuum, activating both of his/her languages, but the base language is more activated than the guest language. This unbalanced activation depends on how proficient the bilinguals are in the guest language, their attitude towards code switching and borrowing, and the topic of conversation (e.g., culture-specific topic vs. universal topic). It also depends on whether code switching is more or less common in the conversational setting (Cheng & Howard, 2008).

The second factor involved in the recognition of a guest word is base and guest language activation. It pertains to the proportion of both languages in a context. The fewer guest words there are in a sentence (i.e., less code switch or borrowing density), the less likely bilinguals activate the lexicon of the guest language. The third factor correlates with the syntactic, pragmatic, and semantic constraints of code switching: how plausible the guest word is in the base language context.

A final factor of interest here is properties of the guest word. This factor comprises the same elements that affect word recognition when monolinguals perceive a target word in their native language, including, but not limited to, word frequency, neighborhood size, and uniqueness point (i.e., that point of the word at which only one candidate cohort is left as a possible target word). Moreover, guest word properties can be more or less consistent with the lexicon of the base language. The more the guest word is consistent with the base language, the slower bilinguals are to access it. Spoken guest words are more consistent with the base language when they are less marked with respect to their phonetics (i.e., the quality of their sounds and prosody is legal in the base language inventory), and phonotactics (i.e., the structure of their sounds and syllables is also legitimate in the base language). In both a gating task and a naming experiment, Li (1996) noticed an interaction between properties of the guest word and the base language effect.[13] The results confirmed that Chinese/English bilinguals recognize English guest words that are phonotactically illegal in Chinese, such as *slash*, more easily than English guest words that are phonotactically legal in Chinese (e.g., initial CV words, such as *lead*). The researcher also found that Chinese/English bilinguals recognize English code switches (i.e., English words pronounced with accuracy) earlier and more easily than English borrowings (i.e., English words pronounced with a Chinese accent). All of this suggests that the assimilative effect of the base language depends on the degree of phonological and phonotactic overlap between the guest word and the base language

---

[13] In a gating task, the participant identifies the target word through listening to successively longer fragments of that word. Grosjean (1988) reported French base language effects on English guest words through an earlier gating experiment.

lexicon. However, the author found that the effect of the Chinese base language on the English guest words stops ahead of time before the offset of the guest word; accordingly, no switching cost appears.[14]

In sum, Grosjean proposes that the non-target language varies in its degree of activation depending on some complex factors, and under certain circumstances, word recognition can be language selective, where a delay in the recognition of code switches is possible. The spoken base language effect is consistent with the BIMOLA (Grosjean, 1997, 2008; Léwy & Grosjean, in preparation). According to the BIMOLA, phoneme and word levels of each language are independently organized in two subsets, and interdependently enclosed in a larger set. When a bilingual listens to another bilingual, he/she activates both language networks (subsets), but since the acoustic waves match the network of the base language more than the network of the other language, the base language becomes more activated. This, in turn, results in top-down feedback from the base language to the guest word. Accordingly, a switching cost might occur if the guest word is highly similar to the lexicon of the base language in terms of its segments and phonotactics, and/or the bilingual listener is in his/her monolingual or semi-monolingual language mode.

Results about whether language switching can have an apparent cost are inconclusive. Additionally, the question of whether there is a bias in favor of the lexicon of the carrier sentence when a guest word is processed is still open to further investigation. My paper examined whether literate speakers of Arabic experience a

---

[14] According to Li (1996), the amount of time that the Chinese/English bilinguals required to identify the English guest words of one-to-two syllables is similar to the amount of time that English monolinguals usually need to identify English words of the same number of syllables.

lexical switching cost by measuring their RTs on target words in language/variety switch and non-switch conditions. On the basis of Grosjean’s proposal, we hypothesized:

H2.2: Literate speakers of JCA process Arabic target words in the context of the same variety of Arabic (language-consistent condition) faster than Arabic target words in the context of the other variety of Arabic (code switching condition).

Some current views suggest an important role for the dominant language (i.e., the language that is used more frequently) in code switching as another source of processing bias. Heredia and Brown (2004) demonstrate that bilinguals code switch from the dominant language to the less dominant language more frequently than they code switch from the less dominant language to the more dominant language. Heredia and Brown propose that this asymmetrical use of code switching may result in an asymmetrical pattern of switching cost. This suggests that bilinguals retrieve code-switched words of the less dominant language faster than they retrieve code-switched words of their dominant language. On the basis of the unbalanced frequency of the two types of code switching, my research introduced the following hypothesis:

H3.2: Literate speakers of JCA retrieve JCA code switches in the context of MSA sentences more slowly than MSA code switches in the context of JCA sentences.

Bilingual research suggests three possible sources that could make bilingual process one of their two languages faster than the other language: the relative level of proficiency in both languages (i.e., L1 vs. L2), the context of interaction (same language context vs. code switch context) and the direction of code switching (switching to the non-dominant language vs. switching to the dominant language). My study examined how diglossic speakers of Arabic react to the two varieties of Arabic compared to bilingual speakers of independent languages. This was the first study to address Arabic word recognition in meaningful sentential contexts rather than as isolated words. It was also the first attempt to test the effect of diglossic code switching on the recognition of Arabic words.

2.3 Method

2.3.1 Participants

Forty Arabic-speaking students, 24 males and 16 females, participated in the lexical decision experiment. All of the participants were native speakers of North Jordanian Arabic, a local form of rural Jordanian. The participants were born and lived in Jordan, where they had completed at least 12 years of formal education in Arabic. They were recruited from Yarmouk University, located in the north of Jordan, by posting flyers on the university bulletin boards. None of the participants reported any hearing impairment or using hearing aids, nor did they participate in any of the other three experiments presented in this thesis. The experiment procedures were approved by the Board of Ethics for Human Research at the University of Manitoba, the author's affiliation.[15] The subjects were paid 5 JD ($10 CAD) for their participation.

2.3.2 Materials and Design

The critical targets of the present study were 44 words and 44 pseudowords. Half of the words were Modern Standard Arabic while the other half were Jordanian colloquial Arabic. I selected the JCA target words from *muʕdʒam ʔalfaaðˤ ʔalħayaah lʕaamyah fi lʔurdun* (Dictionary of the Everyday Language in Jordan), and the MSA target words

---

[15] The protocol reference number from the Joint-Faculty Research Ethics Board is *J2013:058*

from *Al-Mawrid Dictionary* (Baalbaki, 2008).[16] Every target word has a translation equivalent in the other variety of Arabic.[17] However, none of the chosen words has a cognate, near cognate, or a shared root in the other variety of Arabic, as previous research argues that bilinguals process cognates and near cognates faster than non-cognates (Duych et al., 2007; Schwartz & Kroll, 2006; Titone et al., 2011). In selecting the target words, I kept their semantic relatedness to the minimum, as research suggests that semantically related words prime each other both within a language (e.g., Meyer & Schvaneveldt, 1971; Sánchez-Casas et al, 2006) and across languages (e.g., Guasch et al., 2011; Perea et al., 2008; Schoonbaert et al., 2009). Additionally, none of the word stimuli contained language-specific phonological or phonotactic cues, as they were selected from the subset of the vocabularies where the sounds and sound structures of JCA and MSA overlap. Thus, the target words of one variety were phonologically legal in the other variety.

As there is no source of word frequency counts in JCA, I relied on my intuition as a literate native speaker of Jordanian Arabic to choose JCA and MSA words of comparable frequency. I estimated the frequency of the candidate words by asking another group of the same student population, who had not participated in the lexical decision experiment, to sign into a predesigned online questionnaire. The questionnaire asked the participants to listen to the JCA and MSA words and judge how frequent they

---

[16] Dictionary of the Everyday Language in Jordan includes words that are common to Jordanian society as a whole rather than words that are specific to some-local dialect(s).

[17] Some of the JCA critical items used in this experiment, and in the next two experiments, are words borrowed from English long time ago and have become an integrated part of the JCA lexicon (e.g., taksi, balkoneh). Results did not show RTs difference between these English origin words and the rest of the JCA critical items. These words are similar to many JCA words in their phonological structure (i.e., have high word likeness and share large neighbourhood density with many Arabic words).

are, in their usual contexts of use, on a Likert-scale between 1 (least frequent) and 7 (most frequent). The mean ratings of word frequency were used as covariates in the statistical analyses of the lexical decision experiment.

The pseudowords were constructed by changing one or two phonemes of real JCA and MSA words. The pseudowords were also phonologically and phonotactically legal in both varieties of Arabic but absent as forms in either variety. The three types of stimuli (MSA words, JCA words, and pseudowords) matched in respect to the number of phonemes and syllables.

Every word stimulus appeared in the final position of a carrier sentence. The carrier sentences were composed with the intention of keeping their semantic constraint low to reduce any possible effect of cloze probability (i.e., to reduce the probability of the target word completing the carrier sentence frame so that the listener cannot anticipate what the final word is in advance). The same neutral carrier sentence ended with a target word of either the same or the other variety of Arabic. The main point of having the same carrier sentences in both conditions was to reduce the amount of uncontrolled variation. The same word stimulus appeared in two different carrier sentences, one in JCA and the other in MSA, to help diminish any lurking differences in cloze probabilities. The two types of target words (JCA and MSA) and the two types of carrier sentences (switched and non-switched) created a two-factor design (see the examples in Table 2.1).

Table 2.1: *Sample stimuli, JCA and MSA words in language switching and non-switching conditions. The language of the carrier sentence appears to the left of the arrow while the language of target words is on its right side. The MSA language is marked in bold.*

JCA→ JCA

yaa reit	yaa	sitii	wagafa-t	ʃ-ʃaɣlih	ʕalaa	ʃwayit	nasˁaaħah
wishing	VOCATION	madam	Standing-	DEF-thing	on	little	obesity

I wish it had been just a matter of obesity.

JCA→ MSA

yaa reit	yaa	sitii	wagafa-t	ʃ-ʃaɣlih	ʕalaa	ʃwayit	***zukaam***
wishing	VOCATION	madam	Standing-F	DEF-thing	on	little	cold

I wish it had been just a matter of cold.

MSA→ MSA

laa	**ʔa-ʕtaqid**	**bi-ʔanna**	**haaðaa**	**l-ʔamr**	**sa-yuʔadyy**	**ʔilaa**	**zukaam**
No	1sg.pres-think	in-that	this.M	Def-matter	Fut-3.M.pres-lead	to	cold

I don’t think that matter will cause a cold.

MSA→ JCA

laa	**ʔa-ʕtaqid**	**bi-ʔanna**	**haaðaa**	**l-ʔamr**	**sa-yu-ʔadyy**	**ʔilaa**	nasˁaaħah
No	1sg.pres-think	in-that	this.M	Def-matter	Fut-3.M.pres-lead	to	obesity

I don’t think that matter will cause obesity.

A total set of 176 sentence stimuli (44 JCA→JCA, 44 MSA→JCA, 44 MSA→MSA and 44 JCA→MSA) were rotated across four counterbalanced lists to ensure that no participant would listen to the same carrier sentence or target word more than once. Ten participants were assigned to each list. I also embedded the 44 pseudowords equally in JCA and MSA carrier sentences and added them to each list. Thus, each participant listened to 88 sentence stimuli in total. The JCA and MSA carrier sentences were comparable in terms of the number of syllables and words. The code switch sentences were consistent with the syntactic principles that govern code switching between colloquial and Standard Arabic (Eid, 1988). I also asked two educated native speakers of Jordanian Arabic to check the plausibility of the tested sentences.

2.3.3 Procedure and Apparatus

The procedure and apparatus of this experiment were similar to Chapter One. However, the participants listened to sentences rather than to single stimuli. The participants were asked to respond as quickly and accurately as possible by pressing the YES key if the sentence ended with a real word, either JCA or MSA, and the NO key if it did not. Participants pressed the key marked ‘YES’ with the index of their dominant hand, and the key marked ‘NO’ with the index of the other hand. I determined the onset times and durations of the target words aurally and visually using a waveform editor. A participant could not listen to the next sentence until he/she pressed one of the two buttons. After every response, there was a 2000 ms interval of silence combined with a hash mark presented in the middle of the computer screen. The computer randomized every session block so that none of the participants had exactly the same order of sentences. Participants began with a block of 10 practice trials. The estimated time for each session was approximately 15-20 minutes. The stimuli were recorded in a sound-attenuated booth with a PMD660 Marantz digital voice-recorder. The speaker was a male native speaker of Jordanian Arabic (not the author), who spoke the same local dialect of the population and had received his school education in the Arabic language.

2.4 Results

No items or subjects were excluded as a result of excessive error rates. Twenty-six incorrect responses (1.477% of the observations) were removed. Visual inspection of the distribution of RTs showed that RTs shorter than 500 ms and longer than 1500 ms were clearly outliers, resulting in the removal of 37 observations (2.134% of the correct responses).[18]

The overall results, depicted in Figure 2.1, suggest that RTs were longer in the switching condition compared to the non-switching condition, indicating a switching cost in both directions (MSA→JCA and JCA→MSA). They also illustrate that RTs between the JCA and the MSA target words were relatively comparable in the non-switching condition (i.e., JCA→JCA and MSA→MSA). However, the participants took longer to respond to the JCA words in the language switching condition (i.e., MSA→JCA) relative to the MSA target words in the same condition (i.e., JCA→MSA).

---

[18] Some studies on word recognition consider response accuracy as another dependent variable in their analyses. This was not possible in the present investigation, as the participants were near ceiling with about 98.5% of correct responses.

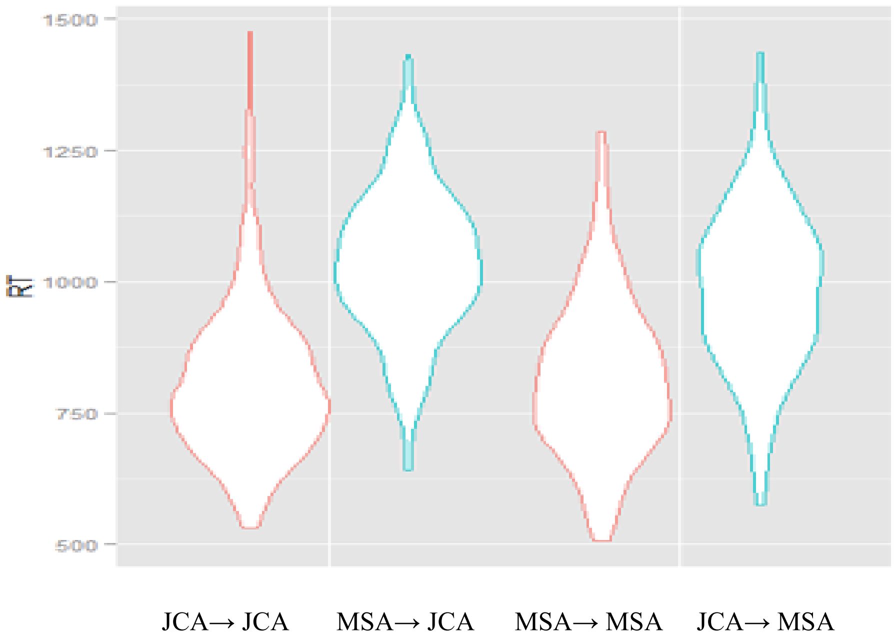


Figure 2.1: RT to words by switching and language of the target word

To validate the pattern of results, I created a mixed-effects model in R, version 3.2.3 with *lme4* (Bates et al., 2016) and *lmerTest* (Kuznetsova et al., 2016) packages. The mixed-effects model incorporated reaction time, measured in milliseconds, as the dependent variable, with language switching condition (coded as switch versus non-switch) and the language of the target word (JCA versus MSA) as independent variables. Additionally, the participant, the carrier item, and target item, were incorporated into the model as random effects. The subjective rating of word frequency was included in the model as a control measure. Table 2.2 summarizes the results of the fitted model with the estimated coefficient (β) in the second column, and their standard errors (SE), *t*-values and *p*-values in the subsequent columns. All effects with $t > 2$ or $-2 > t$ ($p < 0.05$) were treated as statistically significant.

I ran and compared two different models. Initially, I created a reduced model with no interaction between the language of the target word and switching variables, summarized in Table 2.2. The results from this model show that participants took significantly longer time to respond to the switch condition compared to the non-switch condition by, on average, 197 ms. They were also significantly slower to respond to JCA than to MSA target words by an average of 24 ms. Statistical analysis also revealed a significant effect for word frequency, which was included in this model as a control measure, $\beta = -20.34$ ms.

Table 2.2: *Summary of the simple mixed-effects model (the model without interactions).*

Fixed effect	Estimate	SE	*t*-value	*p*-value
Intercepts[6]	923.90	34.54	26.75	< 0.0001
Switching: switch	196.62	6.17	31.89	< 0.0001
Language.of.target.word: MSA	−23.81	6.29	−3.79	0.00016
Frequency	−20.34	6.38	−3.19	0.00174

I also created a full model with an interaction to see whether there is any interdependence between switching and language of the target words. The results from the interaction model show that the average RT to JCA target words in switch sentences (MSA→JCA) was significantly slower than the average RT in the baseline condition of JCA targets in non-switch sentences (JCA→JCA), $\beta = 227.35$ ms, ($t = 12.03$, $p < 0.0001$). There was no significant difference between the MSA and JCA targets in their non-switch conditions (MSA→MSA vs. JCA→JCA), $\beta = 6.86$ ms, ($t = 0.36$, $p = 0.7183$). The interaction between the switching condition and language of target words was nonsignificant although it approached significance, $\beta = -61.39$ ms, ($t = -1.72$, $p =$

0.0928). Figure 2.1 indicates the direction of this possible weak interaction. It suggests that the positive increase in response time moving from the non-switch to the switch condition is relatively greater in cases where the language of the target word is JCA (i.e., MSA→JCA sentences) as compared with MSA (i.e., JCA→MSA sentences). The frequency effect remains significant, $\beta = -20.38$ ms, (t = -3.23, p = 0.0015). To summarize, the full mixed-effects model shows that only the switching condition, but neither the (simple) effect of target word language nor the interaction, affects the reaction times.

To see how well the simple model without the interaction fits the data relative to the complex model, we compare them directly against each other by giving them both to the ANOVA () function. The *p*-value of the chi-square was 0.083, which was nonsignificant at the 0.05 level. This means the more complex model with the interaction was not significantly better than the simpler model without the interaction. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are more appropriate measures of how well a model fits its data. The model with a lower AIC or BIC fits the data better. Both the AIC and the BIC penalize a model for the amount that the data disagree with its prediction and the added complexity. The only difference between the AIC and the BIC is how much they penalize a model for its complexity (Wagenmakers, 2007): the BIC probably penalizes complexity a little too much (more conservative) while the AIC probably does not penalize complexity enough (less conservative). Results of the comparison reveal that the BIC value of the no-interaction model (= 21509) is smaller than that of the interaction model (= 21513). Thus, the BIC prefers the simple model (the one without the interaction) to the complex model (the one with the interaction). The AIC of the interaction model (= 21465) is slightly smaller than that of the no-interaction model (= 21466). This means the AIC prefers the more complex model (the one with the interaction) to the simple model. I would be justified in saying that the model without the interaction is better than the model with the interaction. However, there might be a chance of an interaction that the experiment did not have enough power to detect, but it would require more research to demonstrate it.

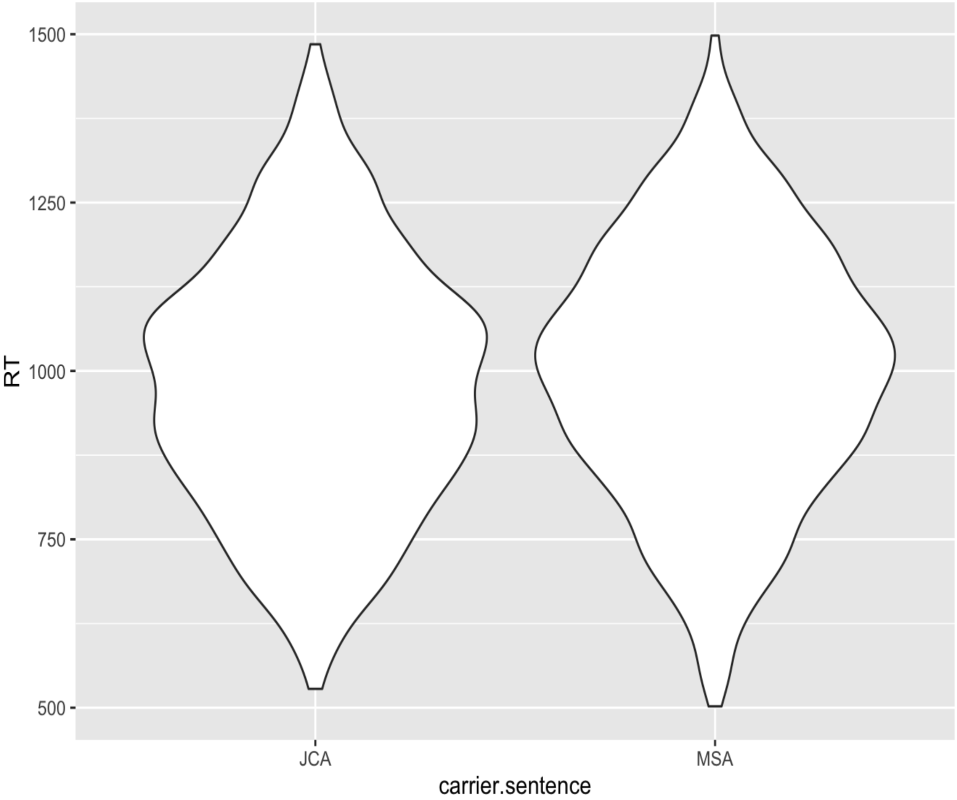


Figure 2.2: RT to pseudowords by the base language.

The analysis of listeners' RTs to the pseudowords did not show significant difference between the nonwords in JCA carrier sentences and the nonwords in MSA carrier sentences. As listeners' pseudoword decision does not seem to differ greatly depending on the base language, there is no evidence that listeners are more likely to process pseudowords as being only MSA or only JCA.

In summary, the results were inconsistent with the prediction of *H1.2* (to repeat: literate speakers of Jordanian Arabic react to MSA words significantly slower than they react to JCA words). On the contrary of the hypothesis, RTs to the MSA words were significantly faster than RTs to the JCA words by 24 ms. However, the results supported *H2.2*, RTs to the target words in the context of the different variety of Arabic are significantly slower than RTs to the target words in the context of the same variety of Arabic, showing a 197 ms switching cost. The statistical analysis did not provide enough support for *H3.2*, which states that literate speakers of Jordanian Arabic retrieve JCA code switches in the context of MSA sentences slower than they retrieve MSA code switches in the context of JCA sentences.

2.5 Discussion

The present study measured response times to JCA and MSA target words in both language switching and non-switching conditions to investigate the status of the two varieties of Arabic and the possibility of switching cost for the diglossic speakers of Arabic. The data showed that native speakers of Arabic judge the target words in non-switching conditions faster than the target words in switching conditions. The data also revealed that native speakers of Arabic judge MSA target words faster than JCA target words. These findings suggest that, at least in the context of this experiment, speakers of Arabic access MSA and non-switched words more easily than CA and switched words.

The findings of this paper differ from earlier studies that found no lexical switching costs in isolated word recognition (Caramazza & Brones, 1980; Thomas &

Allport, 2000; von Studnitz & Green, 2002). This suggests that alternating between individual words of both languages may not be the appropriate way to examine how bilinguals process code switches. When the experiment introduces a single individual word to bilinguals, this may not be sufficient, as a linguistic cue, to reduce the activation of the other language lexicon for the next word. Thus, words from both languages are likely to compete for selection in both switch and non-switch trials of individual words. The findings of the current study are in consonance with some of the previous studies that found switching costs when examining mixed-language processing in sentential contexts (Proverbio et al., 2004; Soares & Groasjean, 1984, Domenighetti and Caldognetto, 1999, as cited in Grosjean, 2008). The present study proposes that when bilinguals listen to sufficient language cues (i.e., a string of cohesive base language words), they activate the lexicon of the base language more than the lexicon of the other language. In turn, this results in a strong base language effect on guest word processing and a possible switching cost. I find it interesting that switching costs would appear not only in bilingual but also in diglossic mixed-language processing. The present data show that the psychological reality of CA and MSA in the cognitive system of native speakers of Arabic resembles the psychological reality of distinct languages in the cognitive system of bilinguals. This suggests one of two possibilities: either diglossia is a kind of bilingualism in itself (Baetens-Beardsmore, 1986), or the switching cost of bilingualism is confounded with any unexpected general linguistic shifts such as part of speech and gender agreement (Bolte & Connine, 2004). I will return to these possibilities later in the discussion below.

The pattern of results in this paper is clearly different from the previous studies that reported no switching costs in sentential contexts (e.g., Gullifer et al., 2013; Li, 1996;

Schreier, 1998;). These inconsistencies could be attributed to some discrepancies in the methods used. Gullifer and his colleagues (2013) examined language switching cost across sentences by asking a group of Spanish/English bilinguals to name a visually marked target word embedded in the middle of a same language sentence. The language of the sentence switched every two sentences. This methodology is problematic for two reasons. First, the switching cost might have been eliminated by virtue of the same-language words presented prior to the target word. In other words, the bilinguals already activated the lexicon of the target word to process the beginning of the sentence. Second, language switches were highly predictable, as the language of the sentence context regularly switched after every other sentence.

The differences between studies could also be because of the interactive role of guest word properties and/or the language mode of bilinguals (Grosjean, 1995; 1997; 2008). To start with the properties of the guest words, the code switches I used in this study were more overlapped with the base language compared to those used in other similar studies. For example, the code switches in Schreier (1998) have little to overlap with the base language because, according to the author, they become language-specific early after their beginning. Guest words can also be less consistent with the base language at the suprasegmental level such as prosody and tone. For example, in Li's (1996) experiments, the tonal nature of Chinese could be the reason for the early identification of the English guest words embedded in a Chinese sentence. That is, the Chinese/English speaker might have pronounced the English code switches and borrowed words with few or no tones, which made them perceptually salient compared to the Chinese base language. In the present investigation, the guest words had very little to distinguish them from base language words, which is in favor of late recognition (even after their offsets). Recall that the base language pushes the listener towards a base language word and in this case. Because there was so little countering at the level of the stimulus, that push was probably quite strong. In addition, the participants had no constraining semantic context in the carrier sentence, which could have speeded recognition of the guest word if the context had favored a guest language word.

Additionally, it could be that the experimental environment and the participants both favored a semi-monolingual mode in the sense that code switching is not frequent in the environment where I ran the current study (Grosjean, personal communication, November 12, 2013). One possible reason for the potential semi-monolingual mode of the participants is that although code switching with single word insertion is adequate in Arabic diglossia, it could be less common than other types of code switching such as alternation (i.e., intersentential code switching) and congruent lexicalization. In congruent lexicalization, “the grammatical structure is shared by languages A and B, and words from both languages A and B are inserted more or less randomly” (Muysken, 2000, p.8). This means that congruent lexicalization code switching occurs within several levels of the sentence and both languages contribute to its grammatical structure. Congruent lexicalization is very common in the speech of educated Arabs, especially in semi-formal registers, where researchers refer to it as Educated Spoken Arabic (Badawi, 1973; El-Hassan, 1978a, 1978b; Meiseles; 1980; Moshref; 2009; Owens, 2001). When speakers of Arabic listen to a string of words in pure MSA or CA without any lexical, phonological, or morphological indications that words from the other variety might be introduced, as the case with the stimuli of the present research, they put themselves in a semi-monolingual mode. Thus, different norms of code-switching can result in different expectations and different processing durations.

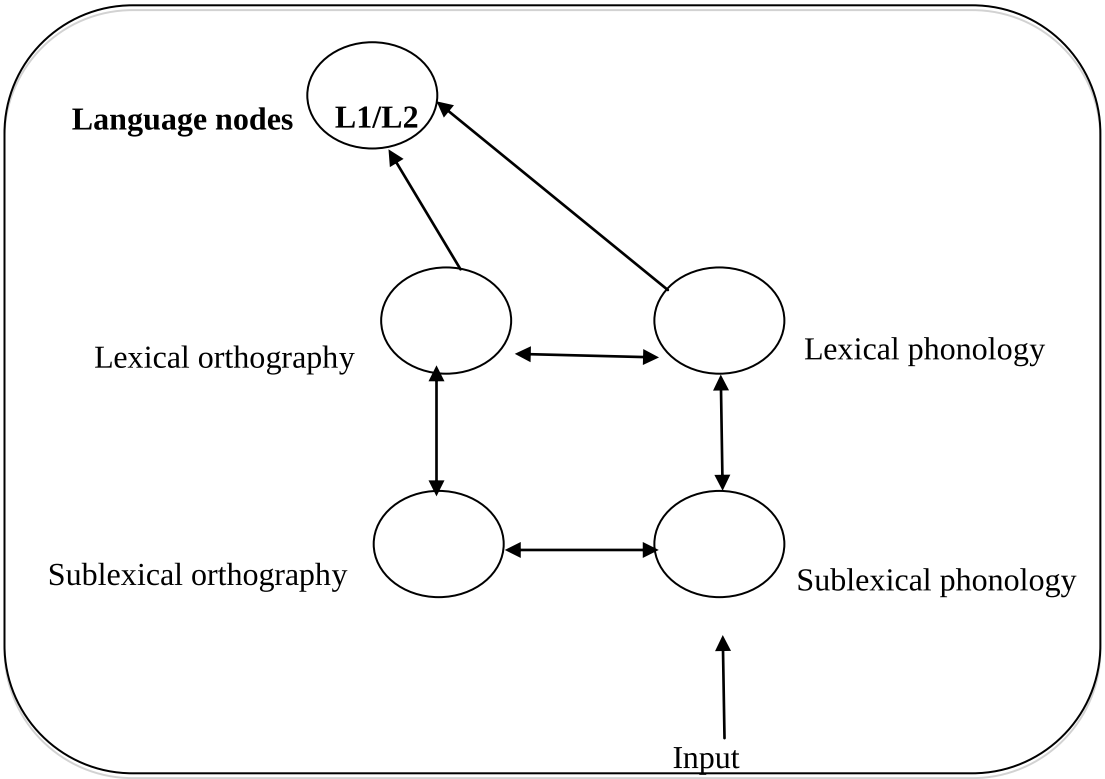


Figure 2.3: *The BIA+ model of bilingual word recognition (Dijkstra & Van Heuven, 2002, p.182). The arrows represent excitatory activation, whereas lines ended with bold circles depict inhibition.*

We can evaluate the BIA+ and BIMOLA models of bilingual word recognition in light of the current findings. The lexical switching cost found in this work is inconsistent with what the BI+ model predicts. In this model, language processing is always nonselective, and there should be no lexical switching cost in any context. Bilinguals activate words of both languages in parallel. Cohorts from both L1 and L2 compete with each other in one integrated lexicon through lateral inhibition until the target word reaches the recognition threshold. Even when the outcome of every cycle of activation (i.e., the activation flow between the layers) is a word from the same lexicon, the bottom-up feed-forward process remains the same and the competition between L1 and L2 words cannot be biased by any top-down feedback from the language node level, as Figure 2.3 depicts.

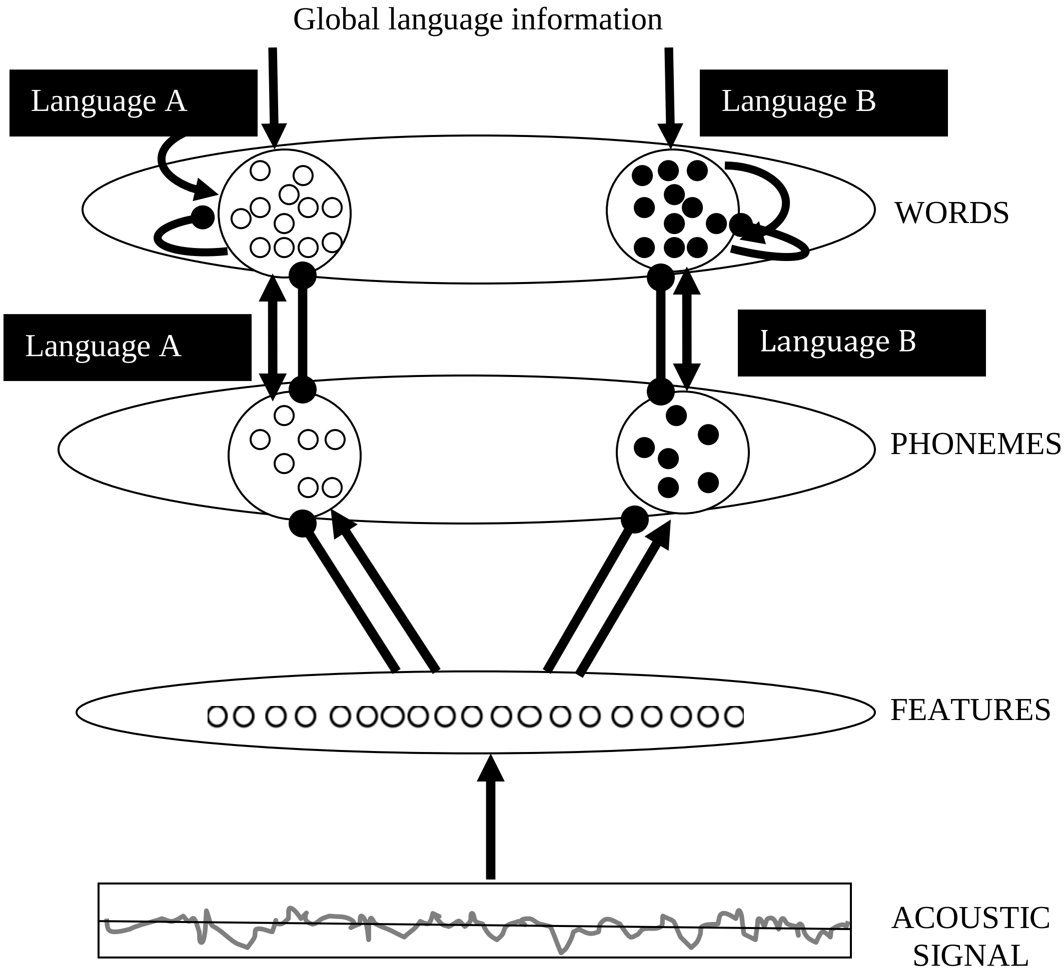


Figure 2.4: *The bilingual model of lexical access (BIMOLA) (Léwy & Grosjean, in prep, as depicted in Thomas & Van Heuven, 2005, p. 211). A model of bilingual speech perception.*

The present findings are consistent with the BIMOLA model of word recognition. The BIMOLA proposes that processing selectivity is a matter of degree, which depends on some complex factors. The BIMOLA argues that bilinguals activate words from both languages in parallel. However, as Figure 2.4 shows, the BIMOLA has language top-down feedback and separate networks (e.g., phonemes or words) for each language. This means L1 words only compete with other L1 words and L2 words only compete with other L2 cohorts. When a bilingual listens to another bilingual, he/she activates both language networks. Word after word, the base language network becomes more activated because there are more similarities between the input and the base language words. This accounts for the base language effect. Bilinguals can increase or decrease the level of activation for the other language (i.e., the non-base language) depending on the proportion of code switches and borrowings in a context. The top-down effect of the base language on the guest word can be weak or strong. If it is weak, the delay on the guest word recognition is made up quickly by the virtue of specific guest language cues (e.g., phonemes, syllable structures, prosody). These cues may attenuate or even eliminate the switching cost. However, when there are not enough language-specific cues boosting the guest language before the offset of the guest word, a lexical switching cost occurs. Furthermore, bilingual listeners may activate one language system over the other on the basis of their language mode. The same bilingual can be in a monolingual mode, expecting no switching to the other language, in some conversational settings; and in a bilingual mode, expecting frequent language switching, in other settings (Cheng & Howard, 2008).

The main architectural differences between the BIA+ and BIMOLA are the language feedback and number of lexicons. The findings of this paper provide evidence for feedback from language identifiers, which goes with the BIMOLA but against the BIA+. If we assume that bilinguals store their languages in two different lexicons, and the switching cost occurs because the higher linguistic information activates one of the two lexicons over the other, then we can propose that JCA and MSA are two different languages. Nevertheless, if we assume that switching cost does not differ from the many other cases of word recognition that clearly operate within a single lexicon and being affected by top-down expectations (e.g., part-of-speech, gender, initial phoneme; Bolte & Connine, 2004; Cole & Segui, 1994), then the switching cost will just be a matter of expectation and not only confined to bilingualism. If this is the case, then BIA+ is the appropriate model, but it should be revised to allow bidirectional activations and inhibitions between the language node and the lexical phonology. To test these hypotheses, we need to know whether switching cost occurs in different registers of the same language such as English formal words in the context of English informal sentences. If it does, this may suggest that switching cost is a general phenomenon that could take place within a single language as well as between two distinct languages. Moreover, future electrophysiological research could measure the event-related potential (ERP; a procedure that measures the electrical activity of the brain over time using electrodes placed on the scalp) of native speakers of Arabic when they listen to JCA and MSA targets in language switch and non-switch carrier sentences. This procedure could determine whether the timing difference in the brain responses takes place in that part of the brain responsible for processing any stimuli or that part responsible for processing the linguistic stimuli in particular.

The guest words used in this experiment are code switches rather than borrowed lexical items as native speakers of Arabic acquire these CA and MSA words at two different stages of their lives and usually use them in different contexts. Future research may compare these results with language borrowing: with the CA and MSA lexical items frequently used in the context of the other variety of Arabic. Diglossic speakers of Arabic are expected to demonstrate less or no switching cost with lexical borrowing compared to code switching. One possible way to distinguish typical MSA and CA words from the words that are also integrated into the other lexicon (i.e., borrowed words) is by asking diglossic speakers of Arabic to judge how frequently a guest word occurs in the context of the other variety.

Now, I will turn to the second important finding: unexpectedly, the participants accessed the lexicon of MSA faster than the lexicon of JCA. This result does not concur with the hypothesis that speakers of Arabic access CA words faster than MSA words because MSA is an L2 for literate native speakers of Arabic (Ibrahim & Aharon-Peretz, 2005). Neither does it replicate Boudelaa and Marslen-Wilson's (2013) findings that native speakers of Arabic access CA words as fast as they access MSA words because the two varieties of Arabic have equal psychological status and serve two distinct sociolinguistic and communicative functions. The important issue that we need to explain is why the participants reacted to MSA significantly faster than they reacted to JCA. The difference between the lexical information of MSA and JCA could be responsible for this bias. The lexicon of MSA has an entrenched orthographic representation relative to the JCA lexicon. Accordingly, speakers of Arabic have less experience in thinking of word boundaries in CA because it is not the language of literacy for them. They are accustomed to seeing spaces between MSA word boundaries, as MSA is far more common than CA in reading and writing. This makes literate native speakers of Arabic identify and judge the final word of a sentence easier when it is a MSA word while thinking slightly longer about what the last word of a sentence might be when it is in CA. This is consistent with the assumption that literacy facilitates metalinguistic tasks such as lexical segmentation or word boundary identification (Petersson, Reis, Askelöf, Castro-Caldas, & Ingvar, 2000). Some aural context research reported that readers can easily isolate words from their semantic content, whereas pre-readers and illiterate adults have difficulty (Allan, 1982; Ehri, 1975; Homer & Olson, 1999; Kurvers, 2002; Kurvers & Uri, 2006).

Allan (1982) investigated lexical segmentation on three groups of children differed in their reading abilities: a nonreadiness group (NR), a readiness group (RR), and a reader group (R). Results showed that the mean segmentation scores were significantly different among the three investigated groups. This supports the inference that children with increasing reading abilities exhibit increasing proficiency in their ability to segment words in both aural and visual sentential contexts. Homer and Olson (1999) asked a group of young children to repeat the final word of a meaningful aural sentence (i.e., a word segmentation task). The authors found that children’s understanding of a word as a unit of speech was predicted by their awareness of words as identifiable units of writing. The children who performed well on the oral task also performed well on the written task. Similar results come from adult illiterates’ marking of word boundaries (Kurvers, 1999, 2002, as cited in Kurvers, & Uri 2006).

In a more recent work, Kurvers and Uri (2006) thoroughly replicated the online methodology of word boundary access (Karmiloff-Smith, Grant, Sims, Jones, & Cockle, 1996) on Dutch and Norwegian monolingual preschool children. The authors reported that the Dutch and Norwegian preschool children scored low yet successful rates of word boundary segmentations (26%) compared to their English monolingual peers, who scored highly successful rates (85%) in the original English study (Karmiloff-Smith, 1996). Kurvers and Uri suggest that the difference between the preschool curricula in England, the Netherlands, and Norway may account for the participants' discrepant performances. The English national preschool curriculum offers indications of formal reading instruction, but there is no formal reading instruction either in the Dutch or in the Norwegian preschool curricula. It is possible that the children in the English study outperformed the Dutch and Norwegian children because they had more experience with written forms. All of this suggests that literacy may play a crucial role in the major changes in children's lexical metalinguistic development.

The present research presents more evidence for how experience with word units in written language (i.e., seeing strings of letters between spaces) might change the online processing of lexical units in sentential contexts. It exhibits that there is more word segmentation ability for the language of literacy compared to the spoken language in aural contexts. This finding is of particular interest because it makes impossible to predict the processing bias just based on the simple views that L2 must be disadvantaged. The orthographic quality of MSA lexicon could be another source of bias making diglossic speakers of Arabic process MSA words faster than dialectal Arabic words in the metalinguistic tasks of sentential contexts. This is because listeners need to identify what the final word is before they decide on its semantic status (i.e., whether it has meaning).

Alternatively, it is possible to assume an interaction between switching and the target word language, but the experiment was not powerful enough to detect this interaction at the $p = 0.05$ level (see the Results section). This possible asymmetrical cost in language switching could be because of the crucial role that the dominant language (i.e., the language that is more frequently used) plays in code switching. According to Heredia and Brown (2004), bilinguals code switch more frequently when they communicate in their dominant language than when they use their non-dominant language. The authors demonstrate that most Spanish/English bilinguals in Spanish-speaking countries code switch from Spanish into English as Spanish is their dominant language. They propose that English words are more expected and, thus, more readily accessible in the context of Spanish while Spanish words are less frequent and less accessible in English contexts. However, the opposite scenario is the norm for Spanish/English bilinguals who live in the United States and use English as their dominant language. Similarly, although MSA→JCA switching is communicatively plausible, it may occur less frequently than CA→MSA switching since CA is the dominant spoken language for native speakers of Arabic. As a result, MSA words become more accessible in the CA context than vice versa. Furthermore, the asymmetrical cost can be related to Arabs' attitude towards MSA. Native speakers of Arabic usually have more positive attitudes toward MSA compared to CA for religious, social, and cultural considerations (Almahmoud, 2013; Feguison, 1959;). Accordingly, their positive attitude toward shifting from CA (a low variety) to MSA (a high variety) may result in a stronger accessibility to MSA lexicon compared to CA lexicon when the shift takes place from MSA to CA.

This investigation could not give a clear answer about whether MSA is an L2 for diglossic speakers of Arabic, probably because of the imposition of word boundary identifications in the employed task. Thus I decided to replicate Ibrahim and Aharon-

Peretz’s semantic priming experiment (2005) on diglossic speakers of JCA. The third chapter will attempt to answer the earlier question posed about whether Ibrahim and Aharon-Peretz’s findings are exceptionally restricted to the Arabic-Hebrew bilinguals because of the sociolinguistic situation of Arabic in Israel, or can be generalized to literate native speakers of Arabic in the rest of the Arab world. In this study, we attained some possible evidence for the effects of literacy on lexical access (word recognition in sentential contexts), the fourth chapter will determine whether there are any similar literacy effects on phoneme access (i.e., sound recognition in the context of carrier words).

Chapter Three

Study III: Modern Standard Arabic in the Minds of Diglossic Speakers of Arabic

3.1 Scope of the Study

The third part of this dissertation reexamines the claim that Modern Standard Arabic (MSA) is a second language for diglossic speakers of Arabic (Ibrahim & Aharon-Peretz, 2005). This study replicates Ibrahim and Aharon-Peretz's (2005) methodology on a different population of Arabic society. The targeted subjects are native speakers of Jordanian Arabic rather than Arabic-Hebrew bilinguals of Israel. The Arabic-Hebrew bilinguals of Israel live on the West Bank of the Jordan River and speak Palestinian colloquial Arabic (PCA) while Jordanians live on its east side and speak Jordanian colloquial Arabic (JCA). Although the two varieties of colloquial Arabic (CA) have some remarkable differences, there is a common impression that Palestinian and Jordanians speak more or less the same dialect, which is classified as the Southern Levantine variety of Arabic. Nonetheless, Arabic-Hebrew bilinguals of Israel were born and grew up in a different sociolinguistic situation, where Hebrew is the dominant language of education and public occasions in Israel (Amara, 2006; Boudelaa & Marslen-Wilson, 2013). The linguistic situation of Arabic diglossia could be more stable in Jordan, as MSA is the only official language used.

In this study, I conducted an auditory translation priming experiment, similar to Ibrahim and Aharon-Peretz's early experiment (2005). The experiment measured participants' reaction times to spoken JCA and MSA target words in semantically related and unrelated conditions. Bilingual research showed that unbalanced bilinguals, who are unequally proficient and practiced in the two languages, react to their L1 words faster than they respond to their L2 words. Similar research also revealed stronger translation priming effects from L1 to L2 but weaker translation priming effects from L2 to L1 (e.g. Jiang, 1999; Schoonbaert et al., 2009; Schoonbaert et al., 2011). Some research did not even report any translation priming effects from L2 to L1, proposing a qualitative difference in how bilinguals access the semantic representations for the vocabulary of both languages (e.g., Finkbeiner et al., 2004; Jiang & Forster, 2001). Sabourin, Brien and Burkholder (2014) concluded that only simultaneous and early sequential English-French bilinguals activate the lexical items of their L1 when they see their L2 translation equivalents. The authors found that late English-French bilinguals do not have shared sematic representations for their L1 and L2 lexical items, even if they are as proficient as early bilinguals. Sabourin and colleagues' results suggest that the age of acquisition is the most relevant factor in L2-L1 semantic activation. The asymmetrical translation priming effect between L1-L2 and L2-L1 is usually introduced as evidence for the assumption that L1 and L2 words do not activate their semantic representations to the same degree (Kroll & Stewart, 1994; Kroll et al., 2002; Talamas et al., 1999). That is, unbalanced bilinguals associate their L2 vocabulary to meaning through loose and weak links, or maybe indirectly through the L1 lexicon while they map their L1 words to their concepts in a robust manner. This study will determine whether JCA vocabulary has a processing advantage over MSA vocabulary. A positive answer to this question suggests that MSA is an L2 for diglossic speakers of Arabic as Ibrahim and Aharon-Peretz (2005) propose. A negative answer to this question supports the proposal that MSA is an exceptional L2 for

Arabic-Hebrew bilinguals of Israel, a thing that cannot be generalized to the rest of the Arab-speaking world (Boudelaa & Marslen-Wilson, 2013).

3.2 Literature Review: Arabic Diglossia in Word Recognition

Ibrahim and Aharon-Peretz (2005) raised the question of whether learning MSA is like learning the formal register of one's native language or it is more like learning an L2. To answer this question, the authors addressed the status of MSA for literate native speakers of Arabic from a psycholinguistic perspective.[19] The researchers conducted two semantic priming experiments on two homogeneous literate groups of Palestinian Arabic-Hebrew bilinguals from Israel. The researchers compared PCA, MSA, and Hebrew words in semantically-related (i.e., translation) and unrelated conditions. In the first priming experiment, the targets were always PCA words and the primes were either in PCA (i.e., intra-language stimuli: PCA-PCA) or in one of the other two languages (i.e., cross-language stimuli: MSA-PCA, Hebrew-PCA). The results showed that Arabic-Hebrew bilinguals reacted to the target words (always PCA) in the semantically-related (i.e., translation) condition faster and more accurately than the target words in the semantically unrelated condition, regardless of the language of the prime. However, the priming effect was three times as large when the presented prime was PCA words (intra-language priming condition) compared to the other two cross-language conditions. Moreover, the priming effect was similar in the two cross-language conditions: equal translation priming effects for PCA targets presented by MSA or Hebrew primes. In the second experiment, the targets were PCA, MSA, or Hebrew words and the primes were always in PCA. The Arabic-Hebrew bilinguals reacted to the PCA targets more accurately and

---

[19] In their early investigation, Ibrahim and Aharon-Peretz (2005) referred to Modern Standard Arabic as Literacy Arabic (LA) and colloquial Arabic as Spoken Arabic (SA).

faster than they reacted to the Hebrew and MSA targets. Moreover, the effect of priming within the PCA language was twice as large as the effect of priming between languages, with no difference in the cross-language priming (PCA-MSA vs. PCA-Hebrew priming). The authors compared the effect of cross-language priming in the two experiments and found it asymmetrical. The magnitude of priming from PCA primes to MSA or Hebrew targets was two times as large as the magnitude of priming from MSA or Hebrew primes to PCA targets, with no significant priming difference when the prime was presented in MSA or Hebrew, and the target was in the other language.

Based on these findings, Ibrahim and Aharon-Peretz concluded that MSA, like Hebrew, is an L2 for native speakers of Arabic. Their conclusions suggest that MSA and Hebrew words are weakly linked to their concepts compared to the PCA words. These results are compatible with the *revised hierarchical model* (RHM; Kroll & Groot, 1997; Kroll & Stewart 1994). The model proposes that associating L2 target words to their meanings is slower and less efficient than associating L1 target words to their meanings, as L2 words are weakly related to their concepts (Kroll, 1993; Kroll & Stewart, 1994). Additionally, an L2 prime triggers weaker semantic expectations, which could mediate activating its L1 translation equivalent at 1000 ms SOA. On the other hand, the slow and inefficient activation of L2 targets can take a better advantage of the more efficient priming induced by L1 words. The model also suggests a developmental shift from weak mapping and lexical mediation to direct mapping between L2 words and their concepts, with increasing fluency (Chen & Leung, 1989; Kroll & Curley, 1988; Tzelgov et al., 1990). In other words, RHM predicts that asymmetric magnitude effects of the semantic priming may be reconciled across individuals or within individuals across time, being determined by the bilingual's proficiency in their second languages.

In a more recent study, Boudelaa and Marslen-Wilson (2013) cast doubt on Ibrahim and Aharon-Peretz's (2005) proposal that MSA is an L2 for literate speakers of Arabic. Boudelaa and Marslen-Wilson conducted two morphological priming experiments: one for MSA words and the other for Spoken Tunisian Arabic (STA). The authors showed that word patterns and roots act as significant linguistic units both in MSA and STA word processing. They reported parallel root/word pattern priming effects in both varieties of Arabic. Furthermore, Boudelaa and Marslen-Wilson found that literate speakers of STA reacted to MSA words as fast as they reacted to STA when they combined the data from the two experiments into a single analysis. On the basis of these results, they concluded that MSA is not an L2 for literate native speakers of STA "despite the differences underlying the two varieties in terms of the productivity of their morphological systems, the age at which they are acquired, and the sociolinguistic context in which they are experienced" (p. 1469). Boudelaa and Marslen-Wilson concluded that STA has no processing advantage over MSA because when children learn an L2 early in their lives (before the critical age of puberty), they gain native-like performance in that language (Birdsong, 1999; Isel et al., 2010), and MSA is a case in point. They also remarked that it is inappropriate to consider any local dialects of Arabic and MSA as L1 and L2, respectively, just because they have different sociolinguistic and functional contexts. That is, MSA and any dialectal Arabic are two varieties with both overlapping and complementary distribution. Finally, Boudelaa and Marslen-Wilson ascribe Ibrahim's findings to the difficult sociolinguistic situation of Palestinian Arabs in

Israel, where Hebrew is the dominant language of education and public occasions. Thus, it is possible that Ibrahim and Aharon-Peretz's (2005) findings reflect some exceptional sociolinguistic situation, which cannot be generalized to the rest of the Arabic-speaking world. This possibility suggests reinvestigating Ibrahim and Aharon-Peretz's priming experiment on a different population of native speakers of Arabic, where the sociolinguistic situation of Arabic is more stable. The present study postulates two opposite general hypotheses. First, MSA is, in fact, an L2 for literate speakers of Arabic and some task or methodological discrepancies might be responsible for the stated controversy between the two reviewed studies. Alternatively, MSA is not an L2 for literate speakers of Arabic, and Arab citizens of Israel are exceptionally unbalanced PCA-MSA bilinguals.

3.3 Research Hypotheses:

Based on the proposal that it is easier to access a target word when its semantic representation has already been activated by a semantically-related (i.e., translation) prime, this study postulates:

> *H1.3: Literate native speakers of Jordanian Arabic respond to spoken words of Arabic in a semantically-related condition faster than they respond to the same target words in a semantically unrelated condition, irrespective of the target language (JCA vs. MSA) or the stimulus type (intra-language vs. cross-language).*

Based on the different ages at which the two varieties of Arabic are acquired/learned and their quantitative experience differences, we hypothesize:

> *H2.3: Literate native speakers of Jordanian Arabic react to the spoken target words of JCA faster than they react to the spoken target words of MSA in both unrelated intra-language and unrelated cross-language conditions.*

Findings of earlier research concluded no lexical switching costs in isolated-word recognition (e.g., Caramazza & Brones, 1980). Accordingly, the present study proposes the following hypothesis:

> *H3.3: Literate native speakers of Jordanian Arabic react to the spoken target words in the unrelated cross-language condition as fast as they react to the same spoken target words in the unrelated intra-language condition, regardless of the target language (JCA vs. MSA).*

As unbalanced bilinguals activate the meanings of their L1 words more strongly than they activate the meanings of their L2 words, this study postulated the following hypothesis:

> *H4.3: The magnitude of semantic priming is larger when the prime is presented in JCA and the target in MSA compared to the magnitude of priming when the prime is presented in MSA, and the target is in JCA. To put it another way,* in cross-*language priming, L1 words are more effective primes than L2 words are.*

This hypothesis is based on the proposal that the concept of the target word is already fully activated when the prime is in L1 while it is partially activated when the prime is presented in L2. In this case, it is the language of the prime that plays the principal role in activating the semantics node and boosting the access to the same semantic node through the target word.

3.4 Method

3.4.1 Participants

Forty native speakers of North Jordanian Arabic participated in this experiment. The participants were undergraduate students from Yarmouk University, located in the north of Jordan. All of the chosen participants had completed at least 12 years of formal education in Modern Standard Arabic. The participants had been recruited via flyers posted in public settings at the university and received monetary compensation for their participation. None of the participants reported any hearing deficit, nor did they take part in the other three experiments.

3.4.2 Materials and Design

A subset of the words used in the present experiment was adapted from the list of words that Ibrahim and Peretz-Aharon (2005) used in their early work.[20] The present experiment reproduced the MSA words that have the same translation equivalents in both PCA and JCA after accommodating their colloquial translations to the phonology of the northern dialect of JCA. There were a total of 12 replicated stimuli. The present experiment included 28 other MSA words with their JCA translation equivalents. All of the related

---

[20] The stimuli are published in Ibrahim (2009).

pairs were chosen to be unique noncognate translations, with clear difference in their phonological forms (e.g., MSA: *ħaqiibah*, JCA: *ʃanteh* ‘suitcase’).

The participants reacted to 80 target words and 80 target pseudowords of Arabic. Half of the target words were MSA words, and the other 40 words were their JCA translation equivalents. Since there were only two tested languages/varieties in the context of this study, I merged Ibrahim and Peretz-Aharon’s (2005) two experiments into one single task. This helps us compare the magnitude of cross-language priming directly within the same participant rather than between two groups of participants. Each MSA and JCA target word was paired with four types of primes: intra-language semantically unrelated prime, intra-language semantically related prime, cross-language semantically unrelated prime, and cross-language semantically related prime. Table 3.1 demonstrates the eight different tested conditions. In the semantically unrelated conditions, the prime words resembled the targets in their frequency, concreteness, and structure. The stimuli were rotated across eight different lists so that none of the participants would listen to the same word (neither as a target nor as a prime) more than once.

Table 3.1: *Sample stimuli, JCA and MSA targets and primes in semantically related and unrelated conditions*

Target	Related Prime		Unrelated Prime	
	JCA	MSA	JCA	MSA
JCA: ʃanteh suitcase	ʃanteh suitcase	ħaqiibah suitcase	ʃubaak window	naafiðah window
MSA: ħaqiibah suitcase	ʃanteh suitcase	ħaqiibah suitcase	ʃubaak window	naafiðah window

Table 3.2 gives the average durations of the prime and target words in both MSA and JCA words. A comparison between the JCA and MSA primes suggests that any potential less effective priming for the MSA words compared to the JCA words might not be just because they have less average durations.

Table 3.2: *Means and standard deviations of the prime and target word durations in milliseconds.*

Prime		Target	
JCA	MSA	JCA	MSA
649 ms (118)	648 ms (116)	660 ms (133)	654 ms (119)

The target pseudowords were derived from real JCA and MSA words by changing one or two of their sounds. The resultant non-words were phonologically legal in one or both varieties of Arabic. The pseudoword followed a priming word of either variety of

Arabic. Thirty judges from the same student population, who did not participate in the main experiment, were aurally presented with a list of the pre-designated MSA words. The judges were asked to sign into an online questionnaire, listen to a set of MSA words, and to suggest for each presented word the best translation in the JCA variety of Arabic. Each word was associated with its best translation equivalent that achieves the greatest consensus among the judges. The lexical database for Modern Standard Arabic (Aralex, Boudelaa & Marslen-Wilson, 2010) provided the tested MSA words with their surface frequencies. The absence of reliable frequency norms for JCA was less problematic in this experiment. Since the JCA words used in this experiment were the best translations for their MSA equivalents, it was assumed that the frequencies of the concepts' translation to MSA and JCA words are similar.[21]

3.4.3 Procedure and Apparatus

The procedure and apparatus of this experiment were similar to Chapter One. However, the participants listened to a pair of items, rather than to a single item, at a time. The participants listened to a prime word followed by a target word or pseudoword. The stimulus onset asynchrony (SOA; the time allotted between the beginning of the prime and the target onset) was 1000 ms. The experiment asked the participants to judge whether the second string of sounds (i.e., the target item) was a real or nonsense word by pressing a YES or NO key. The participants were asked to respond to the target as quickly and accurately as possible. The dominant hand always made the YES response.

---

[21] Yet, I should admit that this method of frequency measures is not optimal. Written and spoken languages/varieties tend to deal with different topics and refer to different things. So almost any written-language corpus count is going to overestimate the spoken-language frequency of words like "president" and "mineral" and underestimate the frequency of "cup" and "please". Future research should revisit and address this issue.

The instructions emphasized that the target word could be in JCA or MSA. Reaction times (RTs) and response accuracy were collected via a SONY portable computer PC (CPU 2.40 GHZ) running Windows 7 and E-prime 2.0 presentation software (Psychological Software Tools, Inc., Pittsburgh, PA, USA; http://www.pstnet.com). The computer reported RTs from the onset of the target. There was an interval of 2000 ms separating subject’s last response from the next stimulus. The primes and the targets were presented over headphones at a comfortable sound level. The participants were tested on only one experimental list and began the experiment with a block of 10 practice trials. The stimuli were recorded in a sound-attenuated booth with a PMD660 Marantz digital voice-recorder. The speaker was a male native speaker of Jordanian Arabic (the same voice in the other three experiments of this thesis), who speaks the same local dialect of the population and has received his school education in the Arabic language.

3.5 Results

No items or subjects were excluded as a result of excessive error rates. Forty-eight incorrect responses (3% of the observations) were removed from the latency analysis. Visual inspection of the distribution of reaction times (RTs) by subjects showed that RTs longer than 1500 ms and shorter than 625 ms were proper upper and lower limits between outliers and the other responses. Thirty-five outliers (0.2% of the data) were discarded from the latency analysis.

To test the research hypotheses, I submitted the data to a mixed-effects model using language R, version 3.2.3 with *lme4* (Bates et al., 2016) and *lmerTest* (Kuznetsova et al., 2016) packages. The mixed-effects model treated reaction times (RTs), measured in milliseconds, as the dependent variable. Participants, prime and target items were also incorporated into the model as random effects. Two types of target words, two types of stimulus conditions, and two semantic relatedness conditions created a three-factor design. The predictors were the target language (MSA vs. JCA), the stimulus type (intra-language vs. cross-language), and the semantic relatedness (related vs. unrelated). To reduce any background noise in statistical analysis, the model incorporated word frequency, UP, and word duration as nuisance variables.

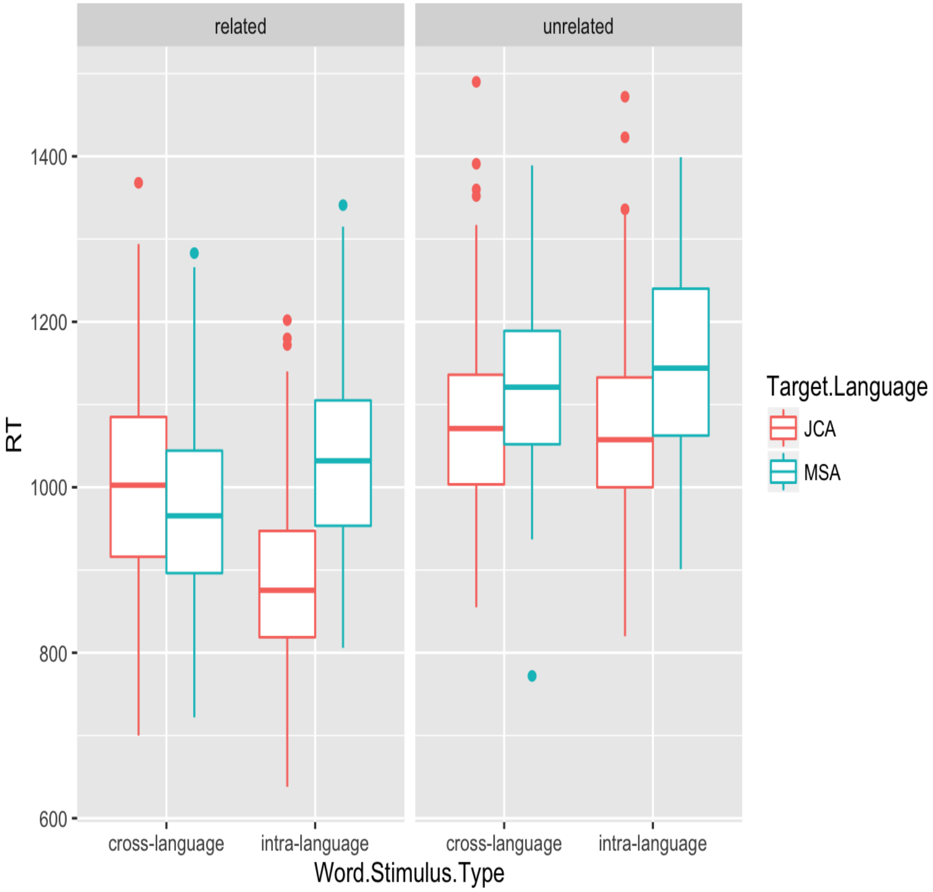


Figure 3.1: *RT by language of the target word, stimulus type, and semantic relatedness*

Figure 3.1 shows that participants' average response to the target words in the semantically-related condition was faster than their average response to the same target words in the semantically unrelated condition. The boxplots also demonstrate that this priming effect was pervasive, depending neither on the language of the primes nor the language of the targets. Inferential statistics, listed in Table 3.3, validated this finding. It revealed that when other factors were held constant (i.e., when both the prime and the target were presented in JCA, with zero frequency, zero UP, and zero word duration), participants' mean RT to the semantically-related targets was significantly ($p < 0.0001$) smaller than their mean RT to the semantically unrelated targets by 183 ms. The nature of the interactions (listed in Table 3.3) and their directions (shown in Figure 3.1) indicate that both the language of the target and the language of the prime depend on the level of semantic relatedness, but not vice versa. These findings are consistent with the first research hypothesis, which predicts semantic priming effects irrespective of stimulus type or target language.

Table 3.3: *Summary of the mixed-effects model for both categorical and continuous variables predicting RTs to target words.*

Fixed effect	Estimate	SE	*t*	*p(t)*
Intercept	982	23.3	42.1	< 0.0001
Relatedness: related	−183	14.0	13.0	< 0.0001
Target language: MSA	77	14.3	5.4	< 0.0001
Stimulus type: cross-language	8	14.0	0.5	0.5721
Target language: MSA*Stimulus type: cross-language	−28	24.6	−1.1	0.2603
Relatedness: related* Target language: MSA	66	19.8	3.3	< 0.0001
Relatedness: related* Stimulus type: cross-language	115	19.8	5.8	< 0.0001
Relatedness: related* Target language: MSA* Stimulus type: cross-language	−161	34.7	−4.6	< 0.0001
Log frequency	−9.3	2.2	−4.1	< 0.0001
UP	0.1	0.03	3.8	0.0003
Target word duration	0.08	0.03	2.3	0.0269

The right panel of Figure 3.1 indicates longer RTs to MSA targets relative to JCA targets in both unrelated intra-language and unrelated cross-language conditions. The same panel shows no average response difference between the unrelated intra-language and the unrelated cross-language conditions, irrespective of the language of the target word. Results of the mixed-effect model revealed that participants significantly ($p <$ 0.0001) slowed their RTs to the MSA targets in the unrelated intra-language condition compared to the JCA targets in the unrelated intra-language condition by, on average, 77 ms. The model yielded no significant difference on average RTs between the intra-language and cross-language condition when the other fixed effects were held constant (i.e., when JCA is the target language of the unrelated condition, with zero frequency, zero UP and zero word duration, $p = 0.6$). The mixed-effects model did not reveal a significant interaction between the language of the target and stimulus type in the semantically unrelated conditions ($p < 0.26$). These results agree with the second hypothesis: in the semantically unrelated condition, native speakers of Jordanian Arabic react to JCA targets faster than they react to MSA targets, no matter in what variety of Arabic the prime is presented. The data also concur with the third research hypothesis: native speakers of Jordanian Arabic react to the target words of Arabic in the unrelated language-consistent condition as fast as they react to the same target words in the unrelated language-inconsistent condition, irrespective of the language of the target.

However, the significant interactions between semantic relatedness and target language, and semantic relatedness and stimulus type suggest that the influences of the target language and stimulus type on RTs were not straightforward, but depended on the level of relatedness. As the Figure 3.1 indicates, the difference between the unrelated intra-language JCA condition (mean = 982) and the related intra-language JCA condition (mean = 799) is greater than the difference between the unrelated intra-language MSA condition (mean = 1059) and the related intra-language MSA condition (mean = 942). The significant two-way interaction between relatedness and the target language, shown in Table 3.3, confirms this difference in the magnitude of semantic priming. Figure 3.1 also depicts that the RTs difference between the unrelated intra-language JCA condition (mean = 982) and the related intra-language JCA condition (mean = 799) is larger than the RTs difference between the unrelated cross-language JCA condition (mean = 1000) and the related cross-language JCA condition (mean = 922).[22] The significant two-way interaction between relatedness and stimuli type validates this difference in the magnitude of priming. Moreover, Figure 3.1 shows that the RTs difference between the unrelated cross-language MSA condition (mean = 1039) and the related cross-language MSA condition (mean = 876) is twice as large as the RTs difference between the unrelated cross-language JCA condition (mean = 1000) and the related cross-language JCA condition (mean = 922). The significant three-way interaction between relatedness, target language and stimulus type ($p < 0.0001$), given in Table 3.3, confirms this difference in the magnitude of semantic priming. The data analysis supports the fourth hypothesis of the research: the priming effect between JCA primes and MSA targets is larger than the priming effect between MSA primes and JCA targets.

---

[22] These numbers can be calculated by adding together the Estimate Coefficients, given in the first column of Table 3.3. The intercept is the average RTs to the basic level (i.e., **the unrelated intra-language JCA condition** when the values of the other continuous variables are zero = **982 ms**). The average RTs to **the related intra-language JCA** = intercept + relatedness: related = 982 – 183 = **799 ms**. Similarly, the average RTs to **the unrelated intra-language MSA condition** = *intercept + target language: MSA* = 982 + 77 = **1059 ms**. The average RTs to **the unrelated cross-language JCA condition** = *intercept + stimulus type: cross-language* = 982 + 8 = **1000 ms**. The mean RTs to **the unrelated cross-language MSA condition**= *intercept + target language: MSA + stimulus type: cross-language + target language: MSA * stimulus type: cross-language* = 982 + 77 + 8 – 28 = **1039**. The average RTs to **the related intra-language MSA condition** = intercept + relatedness: related + target language: MSA + relatedness: related * target language: MSA = 982 –183 + 77 + 66 = **942 ms**. The mean RTs to **the related cross-language JCA condition** = *intercept + relatedness: related + stimulus type: cross-language + relatedness: related * stimulus type: cross-language* = 982 – 183 + 8 + 115 = **922 ms**. The average RTs to **the related cross-language MSA condition** = *intercept + relatedness: related + target language: MSA + stimuli type: cross-language + target language: MSA * stimulus type: cross-language + relatedness: related * target language: MSA + relatedness: related * stimulus type: cross-language + relatedness: related * target language: MSA* stimulus type: cross-language* = 982 – 183 + 77 + 8 – 28 + 66 + 115 – 161= **876**. Now the priming effect between any two conditions can be easily obtained. For example, the priming effect for the cross-language JCA condition= 1000 – 922 = 78, and the priming effect for the cross-language MSA condition = 1039 – 876 = 163. The priming effect difference between JCA and MSA targets in the cross-language condition = 163-78= 85.

3.6 Discussion

This study replicated Ibrahim and Peretz-Aharon’s early translation priming experiment on a different group of Arabic-diglossic speakers. The tested participants were Jordanian rather than Israeli native speakers of Arabic. The priming experiment of this study revealed four main findings. First, literate native speakers of Jordanian Arabic react to Arabic target words in the semantically-related condition faster than they react to the same target words in the semantically unrelated condition. Second, when prime and target words are semantically unrelated, native speakers of Jordanian Arabic respond to JCA targets faster than they react to MSA targets. Third, native speakers of Jordanian Arabic demonstrate no switching cost between pairs of unrelated cross-language words and pairs of unrelated intra-language words. Fourth, JCA words are good primes for MSA targets, but MSA primes are not that good for JCA targets. These results suggest a stronger overall activation/connection between JCA words and their semantics. The bidirectional priming effects for both varieties of Arabic suggest that their translation equivalents share the same semantic network, similar to early bilinguals. The asymmetrical priming effects indicate that JCA lexical items are more integrated into the semantic representations compared to MSA lexical items, similar to sequential bilingualism.

Results of this experiment suggest that literate native speakers of Jordanian Arabic process the vocabulary of JCA and MSA the same way unbalanced bilinguals process the vocabulary of their L1s and L2s in similar experiments (Kroll, 1993; Kroll & Stewart, 1994). For example, Schoonbaert et al. (2009) conducted two masked-priming translation experiments on two similar groups of unbalanced Dutch-English bilinguals. Results of the two experiments showed significant translation priming from L1 to L2 (meisje-girl) and from L2 to L1 (girl-meisje) at two different SOAs (i.e., 250 ms and 100 ms). The translation priming was asymmetrical: there were stronger translations priming effects from L1 to L2 than from L2 to L1.

The present study reproduced results of Ibrahim and Peretz-Aharon’s original work on Palestinian Arabic-Hebrew bilinguals. To determine whether MSA is a second language to Arabic speakers, Ibrahim and Aharon-Peretz (2005) compared semantic priming effects in the auditory lexical decision of spoken Palestinian Arabic, MSA, and Hebrew (an L2 to all participants). Primes were either in MSA or in Hebrew and the targets were in PCA, and vice versa. The priming effects from PCA primes to MSA or Hebrew targets were double the priming effects from MSA or Hebrew primes to PCA targets. Moreover, the priming effect was three times as large when both the prime and the target were PCA compared to MSA or Hebrew primes and no difference between Hebrew and MSA primes was found.

All these results are predicted in light of the *revised hierarchical model* (RHM; Kroll & Groot, 1997; Kroll & Stewart 1994). The model argues that concepts are more strongly linked to their phonological representations in L1 than in L2. If we assume that L2 words can address the semantic system directly but weakly, then activating their semantic representations will be less efficient. As a result, the amount of activation that spreads from an L1 prime to an L2 target is more than the amount of activation that spreads from an L2 prime to an L1 target. In the present study what may cause the difference in the priming effect is the strength of activation from the semantic node to the lexical nodes rather than the speed of building up the word nodes of the primes. Note that, on average, 1060 ms was enough time for the participants to recognize an MSA target, make an executive decision about its lexicality, and physically move their finger to press a button. Accordingly, the one-second SOA (i.e., the time between the onset of the prime and the onset of the target) seems to be enough time to activate the lexical nodes of the MSA primes as robust cohort competitors. The proposed account is consistent with Weber and Cutler's (2004) findings. That study analyzed the proportions of their participants' visual fixation on a target picture, compared to other three-distractor pictures, when they listen to its English word. The researchers observed that, by1000 ms, both their English monolingual speakers and their professional Dutch-English bilinguals looked more often at the target picture relative to the distractors. However, the monolingual had more fixations on the target pictures (95 %) compared to the Dutch-English bilinguals (80%). Weber and Cutler concluded that lexical completion is greater for non-native than for native listeners.

The results suggest that despite native Arabic speakers' intensive daily use of MSA when dealing with written language and spoken language in formal situations, MSA remains as a second language for diglossic speakers of Arabic. Reproducing Ibrahim and Peretz-Aharon's findings on a more stable population of Arabic diglossia casts doubt on Boudelaa and Marslen-Wilson's (2013) proposal that Ibrahim and Aharon-Peretz's (2005) findings may reflect some exceptional sociolinguistic situation that cannot be generalized to the rest of the Arabic-speaking world. The question of why Boudelaa and Marslen-Wilson's findings disagreed with Ibrahim and Aharon-Peretz's results, reproduced in this study, is still open for further investigation. It is possible that some methodological matters were responsible for these inconsistencies.

Psycholinguistic research on bilingualism emphasizes that the speed of access to semantic representation is not limited to language proficiencies and the context of acquisition, but also affected by item characteristics, such as concreteness and cognates (De Groot, 1992a; 1992b; 1995; Van Hell, 1998; Van Hell & De Groot, 1998). For example, Van Hell and De Groot (1998) reported that bilinguals demonstrate stronger priming effects between cross-language concrete nouns and cognates (i.e., words sharing meaning and cluster of sounds in different languages) compared to cross-language abstract nouns and non-cognates. Their results agree with the *distributed feature model of bilingual semantics* (De Groot, 1992a; 1995; Van Hell & De Groot, 1998). This model predicts that the semantic representations for concrete nouns and cognate translations are nearly similar across languages while abstract nouns and non-cognates translations are more distinct. This explanation is based on De Groot and colleague's proposal that both concrete and cognate translation equivalents share more distributed meaning and semantic overlap compared to abstract and non-cognate translation equivalents. In a more recent study, Ibrahim (2006) found greater priming effects when PCA/Hebrew primes and MSA targets were cognate translations than when they were non-cognate translations. However, neither the current study nor Ibrahim and Aharon-Peretz's original work examined the effects of concreteness on cross-language priming of Arabic words. Effects of concreteness can be a topic of further future research, especially if a reliable objective measure for Arabic word concreteness is employed.

A quick comparison between the findings of the present priming experiment and results of the word recognition task in sentential context (Study II) suggests two closing remarks. First, MSA is an L2 for the diglossic speakers of Arabic. Yet, whether MSA words have a processing disadvantage relative to CA depends on the experimental context. While diglossic speakers of Arabic showed processing advantages for JCA words in the context of isolated word recognition (the current experiment), they demonstrated faster access to MSA words relative to JCA words in the sentential context of a metalinguistic task (i.e., Study II). Second, the experimental context can also determine whether diglossic speakers of Arabic experience switching costs when they listen to cross-language stimuli of Arabic compared to intra-language stimuli. Findings of Study II confirm that diglossic speakers of Arabic experience switching costs when an Arabic guest word occurs in a coherent sentential context of the other variety of Arabic. The data from the present study showed not switching cost when the target word simply follows individual isolated words from the other lexicon.

Chapter Four

Study IV: Phoneme Awareness in the Context of Arabic Diglossia

4.1. Introduction

Phoneme awareness refers to the listener's ability to notice, identify, and manipulate the individual sounds of spoken words (Anthony & Francis, 2005; Chard & Dickson, 1999). It is a subset of a more comprehensive term, *phonological awareness*, which includes the ability to identify and manipulate the most fine-grained units of spoken words (i.e., phonemes) as well as larger units of sound sequence such as rimes and syllables (Goswami & Bryant, 1990; Mattingly, 1972). Phoneme awareness is a skill difficult to acquire and developed much later than the rime and syllable levels of phonological awareness (Cossu, Shankweiler, Liberman, Katz, & Tola, 1988; Liberman, Shankweiler, Fischer, & Carter, 1974). According to Liberman et al. (1974), it is easier to perceive and manipulate syllables than phonemes because syllables have peaks of acoustic energy, which help listeners think about them as separate units. In contrast, phonemes lack spontaneous cognitive existence because of the constant effects of co-articulation (i.e., phoneme boundary overlapping).

The study presented in this chapter will investigate the role of Arabic diglossia in phoneme awareness. It probes whether and how lexical information intervenes in phoneme identification. Results from some relevant research indicate that phoneme identification depends not only on information directly derived from the speech signal (i.e., prelexical information) but also affected by the stored knowledge about the lexical information of the word (e.g., Castles, Holmes, Neath, & Kinoshita, 2003; Dijkstra, Roelofs, & Fieuws, 1995; Frauenfelder & Segui, 1989; Hallé, Chereau, & Segui, 2000). This study will examine the role of four lexical characteristics of Arabic in phoneme decision (i.e. phoneme identifying). These lexical properties are word uniqueness point, word frequency, the diglossic status of the word, and the phonological neighborhood density of the word (i.e. its phonological family-size). The present work will also examine the role of Arabic orthographic depth in phoneme awareness. The diglossic status of the Arabic word denotes how common the word is used in listening and speaking compared to reading and writing. In the context of Arabic diglossia, colloquial Arabic (CA) varieties are the dominant languages of verbal communication, whereas Modern Standard Arabic (MSA) is far more common in print. Uniqueness point (UP), also known as the recognition point, is the point of the word of which there is only one possible candidate left in the cohort of the other competing words (Marslen-Wilson, 1987). In a word such as *cappuccino* /kæpətʃi:noʊ/ the uniqueness point is the point at which the listener starts to hear the /i:/ sound because all other lexical competitors are automatically deactivated after this sound.[23] Word frequency indicates how familiar language users are with the word. In the context of this study, it means how often we hear and articulate the word. The phonological neighborhood density refers to how words similar in their phonological representations are clustered or stored in memory. Although there is no perfect procedure of measuring phonological neighborhood, it is usually assessed by the single phoneme change. The phonological family size (i.e., the phonological neighborhood density) of a given word may include the number of words that can be created from that word by adding, deleting, or substituting one phoneme (e.g.,

---

[23] [i:] is the UP for *cappuccino* /kæpətʃi:noʊ/ if we assume that the listener knows the word *capuchin* /kæpətʃin/ ‘the type of monkey’, whose UP is the short vowel [i]. The two forms are identical until the listener hears either the long vowel [i:] or the short vowel [i].

Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, part of the neighborhood for the English word *cat* is *mat*, *at*, *scat*, *rat*, *cap*, *etc*. Words with a high number of neighbors have dense neighborhoods, whereas those with few neighbors have sparse neighborhoods. Orthographic depth specifies the consistency and predictability of grapheme-to-phoneme mappings in a language (Ellis et al., 2004). The English <d> is a shallow letter because it corresponds to one sound, /d/ while the letter <x> is orthographically deep because it represents a combination of two different sounds, /ks/. All Arabic and Hebrew consonant letters are orthographically deep (e.g., Arabic <م>) because they usually correspond to the consonant sound and the following short vowel (/mi/, /mu/, /ma/) except in the vowelized/pointed texts, which are less common in both languages. The current study will scrutinize Arabic phoneme awareness through a phoneme monitoring task, in which participants are asked to press a response button as soon as they recognize a target phoneme in a spoken stimulus (e.g., Cutler, Treiman, & van Ooijen, 2010; Dijkstra et al., 1995; Hallé et al., 2000; Morais, Bertelson, Cary, & Alegria, 1986).

4.2. Theoretical Background and Literature Review

Sections *4.2.1* through *4.2.6* of this study will review some of the previous research on the acquisition and development of phoneme awareness. More specifically, Section *4.2.1* will explain how lexical representations of words may influence phoneme recognition. Sections *4.2.2* through *4.2.4* will focus on how alphabetic literacy and orthography can affect phoneme recognition. Section *4.2.5* will discuss the relationship between spoken vocabulary experience and segmental (i.e., phonemic) restructuring of lexical representations. Section *4.2.6* will demonstrate the role of Arabic diglossia in phoneme awareness.

4.2.1. Phoneme Monitoring and Levels of Representation

Some evidence for lexical effects on tasks that tap prelexical representations such as phoneme decision comes from the processing advantage of word phonemes over nonword phonemes. For example, Dijkstra et al. (1995) reported that phoneme monitoring is faster for the target phonemes in Dutch words than for the same target phonemes in nonwords. The effect increases when the target is positioned after the UP of the carrier words and nonwords.[24] Moreover, Frauenfelder and Segui (1989) found that their French-speaking participants responded faster to a given target phoneme when the bearing word was preceded by a semantically-related word (e.g., *clou* ‘nail’/*pointe* ‘point’, *oreille* ‘ear’/*entendre* ‘hear’) compared to an unrelated word (e.g., *curve* ‘tank’/*pointe* ‘point’, resort ‘spring’/*entendre* ‘hear’. The authors present the semantic priming effect as a piece of evidence for the influence of lexical representation on phoneme decisions in generalized phoneme monitoring tasks (i.e., when the target phoneme is not confined to a specific position of the target-bearing word).

Major models of word recognition agree that word recognition influences the phoneme monitoring task. For example, the *TRACE model* (McClelland & Elman, 1986), the *autonomous race model* (Cutler & Norris1979), and the *Merge model* (Norris, McQueen & Cutler, 2000) all predict that conscious awareness of phonemes depends on nodes that get (part of) their activation from word nodes. However, Models of word recognition disagree on the precise layout of those nodes. In what follows, we will briefly see some views of how word recognition may affect phoneme monitoring.

---

[24] The UP for nonwords is the critical point in time of the spoken item where the listener is more confident that the stimulus is meaningless.

The *TRACE model* (McClelland & Elman, 1986) argues that phoneme metalinguistic tasks such as phoneme identification occur in the phoneme node, and there is top-down feedback from the lexical node to the phonological node. However, proponents of the *autonomous race model* (Cutler & Norris1979) disagree with the whole idea of feedback, insisting on a multiple-outlet mechanism where the linguistic inputs map onto both the prelexical phonemic node and the lexical node in parallel. Both layers can autonomously be used to identify a phoneme. Whichever route reaches an output first determines the response. The race between two processes is faster than either process alone. This, in turn, explains why phoneme decisions are faster in words (i.e., two routes) than in nonwords (i.e., one route). The *Merge model* (Norris, McQueen & Cutler, 2000), like the *race model*, disputes the entire idea of feedback. The *Merge model* argues that there are two different layers of phoneme nodes. The first phoneme node is between the linguistic inputs (i.e., sound, letters) and the lexical node. The second phoneme layer is located after the lexical layer. Conscious metalinguistic tasks such as phoneme monitoring and categorization take place in the second layer, and there is no need for feedback to the first layer. The second layer of phoneme units is not permanent but appears on the surface every time we need it. Despite their mechanism differences, all of the three discussed models argue for lexical effects on phoneme recognition.

The experiment presented in this chapter is not concerned with the specific architecture of each model, but the broad difference between these phoneme-monitoring models, which exploit lexical information, (e.g., Cutler & Norris1979; Norris, McQueen & Cutler, 2000) and the phoneme-monitoring model that depends only on the acoustic-phonetic representation (e.g., Foss & Gernsbacher, 1983). In other words, the present study aims to explore the relationship between phoneme monitoring and lexical access in the context of Arabic diglossia rather than to seek support for one specific model. Behavioral research on lexical access per se (e.g., lexical decision) found significant effects for variables such as UP, word frequency, and word duration. Based on the proposal that phoneme monitoring relies on lexical access, this study predicts similar effects for these variables in phoneme monitoring tasks. It may also hypothesize that phonemes of JCA words will be detected faster than phonemes of MSA words based on the earlier findings that isolated CA words (i.e., L1 words) are accessed faster than isolated MSA words (e.g., L2 words) (Ibrahim & Aharon-Peretz's, 2005; Chapter Three of this thesis).

4.2.2. Alphabet Literacy and the Development of Phoneme Awareness

Spoken language seems to have primacy over written language in individuals' lives because we talk before we read, and many human languages do not have a writing system. Yet, a growing body of research proves that reading skills can influence basic phonological processes (Olson, 1996), including our ability to consciously manipulate and identify individual phonemes (e.g., Chueng & Chen, 2004; Dijkstra et al., 1995; Goswami, 1999; Morais, Cary, Alegria, & Bertelson, 1979; Treiman & Cassar, 1997). Morais et al. (1979) compared two groups of adult Portuguese speakers who differed only in their exposure to alphabet reading to measure their degree of phoneme awareness. The authors discovered that the literate adults could add and delete the target consonants at the beginning of nonwords far better than the adults who remained illiterate. Similar research on children suggests that whereas knowledge of syllables and rimes appears to develop spontaneously at preschool age, knowledge of phonemes appears to develop when children go to school and begin to learn to read in alphabetic orthography (Chueng & Chen, 2004; Goswami, 1999). This sequential development is evident in children of different language backgrounds including English speaking children (Treiman & Zukowski, 1991), Italian children (Cossu et al., 1988), German children (Wimmer, Landerl, & Schneider, 1994), Czech children (Caravolas & Bruck, 1993), Swedish children (Arnqvist, 1992), and Arabic children (Amor & Maad, 2013). In their well-known work, Cheung and colleagues (2001) assert that it is the alphabet literacy rather than literacy, in general, that matters in phoneme awareness. The researchers compared the phoneme awareness of first-grade children in mainland China, who learn the Pinyin (i.e. alphabetic) writing system, to their peers in Hong Kong, who learn the usual logography (non-alphabetic) characters, and English-speaking children. The Hong Kong children scored lower levels of phoneme awareness than the mainland Chinese children, who had learned the Pinyin writing system, and the English-speaking children (For similar findings, see Read, Zhang, Nie, and Ding, (1986); and McBride-Chang et al., (2004)). To conclude, previous research suggests that as long as a language employs an alphabetic writing system, all typically developing (non-dyslexic) children, independent of their language background, develop their phonological awareness from awareness of syllables, rimes, and onsets to awareness of phonemes.

Despite the alphabetic literacy constraint on the development of phoneme awareness, research has shown cross-linguistic differences in phoneme awareness based on how salient/important a given unit in the language is. For example, native speakers of Arabic and Hebrew manipulate consonants in the initial position of a syllable with great difficulty compared to the consonants that occur in the final position of a syllable

(Saiegh-Haddad 2007a, 2007b; Share & Blum, 2005). However, the opposite pattern was found among monolingual English children and adults (Bryant, MacLean, Bradley, & Crossland, 1990; Treiman, 1983; 1988). The stronger cohesiveness for Arabic and Hebrew CV body was partially ascribed to the predominance of the CV unit in the phonologies of Hebrew and Arabic (see Section *4.2.4* below for an alternative satisfactory account) while the stronger cohesiveness for English VC rimes was attributed to the dense rime neighborhoods of simple English words (Saiegh-Haddad, Kogan, & Walters, 2010). Thus, the psychological cohesion of Arabic CV (the body) and English VC (the rime) is related to the peculiar phonological structures in both languages.

In conclusion, although a large body of evidence supports the hypothesis that phoneme awareness develops primarily as the product of alphabetic literacy, this effect should be understood as correlational rather than causational (Castles & Coltheart, 2004). In the context of the present study, it is possible to hypothesize that adult native speakers of JCA identify the target phoneme in MSA words earlier than they do in JCA words. This prediction is based on the fact that MSA is the language of literacy for native speakers of Arabic. This is a contrasting view of that JCA, being the L1, would have more built-in awareness, at least at younger ages. In what follows, we will see other aspects of how alphabetic orthography influences spoken phoneme and word recognition.

4.2.3. Spoken Word/Phoneme Recognition and Orthographic Co-activation

The above section demonstrated how phoneme awareness is, primarily, the product of alphabetic knowledge. This section will focus on how orthographic knowledge may affect the recognition of auditory units of language. Different metalinguistic tasks have shown that alphabetic knowledge can influence both spoken word recognition (e.g., Chéreau, Gaskell, & Dumay, 2007; Ranbom & Connine, 2011; Ziegler & Muneaux, 2007) and phoneme awareness (e.g., Castles et al., 2003; Dijkstra et al., 1995; Frauenfelder, Segui, & Dijkstra, 1990; Hallé et al., 2000). Some research demonstrates that orthographic representation is automatically activated during auditory processing tasks and affects spoken word and phoneme recognition even in the absence of visual information. For example, in a simple lexical decision task, Ziegler and Ferrand (1998) found that their French-speaking participants recognized consistent French words such as *stage*, where the phonological rime can only be spelled as <age> as in *stage*, *rage, cage*, more accurately and faster than inconsistent French words such as *prompt*, where the phonological rime can be spelled in different ways as in *nom*, *prompt*, *ton*, *tronc*, and *long.* Another possibility for orthographic co-activation in speech processing comes from auditory priming tasks (e.g., Chéreau et al., 2007; Taft et al., 2008). Chéreau et al. (2007) examined the effects of orthographic overlap between a prime and the target in English real words and pseudowords. The spoken prime and targets were related to their offsets either phonologically (*scheme-gleam*) or both phonologically and orthographically (*dream-gleam)*. The researchers discovered a significant extra facilitation effect when the primes and targets overlapped in both sounds and letters.

As with word recognition, research has also found consistency effects in phoneme awareness tasks (e.g., Dijkstra, et al., 1995; Frauenfelder et al., 1990). In their study, Frauenfelder and the co-authors found evidence that the French phoneme /k/, which can be spelled in more than one way has higher processing costs than those that have only one possible spelling such as /p/ and /t/. In a similar study, Dijkstra et al. (1995) found that their Dutch speakers took longer to detect the phoneme /k/ in words where /k/ is spelled with its less-dominant (i.e., secondary) spelling <c> relative to the words that spell it with <k>. They also reported that the effect was greater when the target sounds located after the uniqueness point of the word, and monitoring times were faster for words than for non-words.[25] Their findings may suggest two stages for phoneme monitoring. In line with this earlier evidence for orthographic co-activation, Hallé et al. (2000) found that their French-speaking subjects perceived the /p/ sound as /b/ in a word like <absurd> [apsyrd]. All of these results concur with the hypothesis that the orthographic representation of words could be activated in both phoneme monitoring and lexical speech processing.

4.2.4. Orthographic Depth and Phoneme Awareness

The correlation between alphabetic orthography and phoneme awareness is also manifested in the degree of orthographic depth. Transparent (or shallow) orthography has a high degree of consistency between letters and phonemes (e.g., one letter represents one phoneme). In contrast, opaque (or deep) orthography has a high degree of irregularity in letter-phoneme correspondences (e.g., letters often represent more than one phoneme or vice versa). Languages vary in their orthographic depth. For example, Finnish and Serbo-Croatian have extremely shallow (transparent) orthography (i.e., one-to-one correspondences between letters and sounds). Semitic languages such as Arabic and Hebrew are extremely deep (opaque), as diacritic markers that correspond to short vowels, and gemination (i.e. a sequence of two identical sounds) are usually omitted in

---

[25] The general sense here is that right after the UP, the listener becomes more certain what the word is and stars to devote more energy to phoneme detection.

print. Other languages, such as French and Russian, exist between these two ends of the orthographic depth continuum (Seymour, Aro, & Erskine, 2003).

Some research has shown that listeners have more difficulty manipulating and counting phonemes that are not present in a word’s orthographic representation (Bassetti 2006; Escudero & Wanrooij, Castles et al., 2003; 2010; Saiegh-Haddad et al., 2010). In their study, Castles et al. (2003) found that their participants had extreme difficulty in deleting or reversing the phoneme /s/ in English words such as *fix* where the /s/ is not spelled by the letter <s> but rather it is encoded, along with /k/, by the single letter <x>. Similarly, Bassetti (2006) discovered that native English learners of Chinese count one fewer vowel when the vowel is not represented in the Pinyin (i.e. Romanized) spelling of Mandarin words. In a follow-up experiment, Bassetti confirmed that English learners of Chinese segment Mandarin vowels as they are spelled in Mandarin. Additionally, Bassetti revealed that English learners of Chinese interpreted the L2 letters as they are represented in their L1 and suggests that the learners are strongly influenced by L1 letter-phoneme conversion rules. Similarly, Share and Blum (2005), and Saiegh-Haddad (2007) consider orthography as a plausible account of the onset/coda segmentation difference stated earlier in Section *4.2.2.* The researchers ascribe the robust cohesiveness between the onset and the following vowel to the deep orthographic systems of Arabic and Hebrew, which utilize their consonant letter (e.g., Arabic initial letter < نـ >) to encode the consonant and the following short vowel (e.g., /na/, /nu/, /ni/, etc.). Interestingly, since the authors used one-syllable words only, their coda consonants were all word-final, and the final shape of consonant letters (e.g., Arabic final letter <ن ـن >) only ever represents one sound (e.g., /n/). This can explain why it was easier for the children to strip off the consonants when they were word final. This is less of a problem for the Hebrew results since Hebrew has fewer different final shapes (e.g., ן). However, Share and Blum did not present the necessary statistics to show that only a minority of words that do have strange finals is not carrying their results. In conclusion, evidence from phoneme manipulation and counting tasks supports the difficulty of separating letter-phoneme associations once those associations are entrenched and crystalized by learning letters. The current research will investigate the orthographic depth of Arabic in phoneme monitoring, postulating that native speakers of Arabic detect Arabic consonants faster and more accurately than short vowels.

4.2.5. Segmental Restructuring of Lexical Representations: The Effects of Word Frequency and Neighbourhood Density

Aside from alphabetic literacy, Metsala and Walley (1998) introduce the *lexical restructuring model* (LRM) proposing that the emergence of phoneme awareness is a more natural process of language development and reflects developmental changes in the organization of the lexicon. A group of researchers (e.g., Elbro 1996; Fowler 1991; Walley 1993) suggest that phoneme awareness arises as a gradual change from holistic to more explicit, segmental representations of words. This segmental restructuring of spoken words results from vocabulary growth, which increases the pressure for fine-grained segmental representation, and extends gradually from younger to older children to adults. Several longitudinal studies employed gating paradigms (Grosjean, 1980; i.e., individuals listen to sequential parts of a word until the word is presented completely) and found that adult listeners need less speech input to recognize words than older children who, in turn, need less input information than younger children (e.g., Elliott, Hammer, & Evan, 1987; Metsala, 1997; Walley, Michela, & Wood, 1995). These observations support the first claim of the LRM. The model proposes that young children do not pay much attention to the fine-grained segmental representations of words because they know a limited number of words. However, with the gradual increase in the vocabulary size, children start to realize different acoustic overlap among words and structure more segmental representations of the stored words (Metsala & Walley, 1998; Walley, Metsala, & Garlock, 2003). Such segmental representation is beneficial because it facilitates discrimination of similar words (Charles-Luce & Luce, 1990) and supports more efficient articulation (Lindblom, 1992; Studdert-Kennedy, 1984). The positive correlation between vocabulary size and lexical restructuring suggests that phoneme awareness is a gradable rather than all-or-nothing phenomenon.

Another related claim for LRM is that phoneme awareness is word-specific, relying on the lexical characteristics of the word compared to the other vocabulary in the same lexicon. Neighborhood density and word frequency are two major lexical characteristics proposed to be associated with phoneme awareness (Metsala & Walley, 1998; Walley et al., 2003). Since words with dense neighborhoods have many phonologically similar words, LRM hypothesizes more phonemic details in words with dense neighbors compared to words from sparse neighborhoods. According to LRM, words with high neighborhood density should be reacted to more accurately and faster than words with low neighborhood density in phoneme awareness tasks. The model also predicts strong segmental (phonemic) representations for high-frequency words but more holistic representations for low-frequency words. This hypothesis also indicates that familiar words should have better-segmented representations much earlier than unfamiliar words.

Although the model does not specify how our word experience can contribute to lexical restructuring, Hogan, Bowles, Catts, and Storkel, (2011) suggest a “word that is produced by multiple speakers, multiple times would contain more phonemic detail as a result of contrasting tokens of phonetic variations of the same word” (p. 50). Adults can also be more sensitive to phonetic details in familiar and frequent words compared to unfamiliar and less frequent words (White, Yee, Blumstein, & Morgan, 2013).

The hypothesized effects of neighborhood density and word frequency have been tested fairly on children (De Cara & Goswami, 2003; Hogan et al., 2011; Metsala, 1999; Roth, Troia, Worthington, & Handy, 2006; Troia, Roth, & Yeni-Komshian, 1996) and scantily on adults (Ventura, Kolinsky, Fernandes, Querido, & Morais, 2007). For instance, Metsala (1999) found that preschool children tend to delete individual sounds more correctly in a phoneme deletion task when they occur in words of a dense neighborhood. Troia et al. (1996) showed that their kindergarteners and second-grade students blended sounds to form low-frequency words with more effort relative to high-frequency words. However, the advantage of high-frequency words was not replicated in their phonemic segmentation task. Garlock, Walley, and Metsala, (2001) did not find word frequency or neighborhood density effects, neither for children nor for adults, when they asked their participants to delete or add an initial sound. The author suggested floor and ceiling effects as a possible account of their findings: the tasks were very difficult for children or very easy for the adults. In a more recent study, Hogan et al. (2011) reported that second and fourth-grade children’s performance on their phoneme deletion task was better for high-frequency words than for low-frequency words and for words from a dense phonological neighborhood than words from a sparse phonological neighborhood.

To conclude, results of some research are consistent with the prediction of LRM concerning the positive relationship between vocabulary growth and phoneme awareness. These findings concur with the LRM claim that lexical characteristics (i.e., word frequency and neighborhood density) contribute to lexical restructuring. My proposed research will consolidate and extend the findings of the previous research on the role of word familiarity and phonological density in phoneme awareness.

4.2.6. Phoneme Awareness in Arabic Diglossia

Effects of the phonological distance between standard and colloquial Arabic on the acquisition of basic language and literacy skills has recently begun to attract some attention. Saiegh-Haddad (2003; 2004; 2007a) conducted a series of studies to examine the effects of the phonological and lexical differences between MSA (the literary variety) and CA (the spoken variety) on the phonemic performances of kindergarteners and elementary school children. The results showed that the children, who accurately articulated the target phonemes, found it more difficult to isolate the Arabic sounds that exist in literary Arabic than those that exist in both varieties of Arabic. The researcher ascribed the difficulty to the children's deficiency in the phonological representation of MSA words. To rule out any probable production effect, Saiegh-Haddad, Levin, Hende, and Ziv (2011) used a phoneme recognition task, which does not require any phonological production, and found that five-year-old kindergarteners recognize MSA phonemes less efficiently than CA phonemes in their carrier words. Thus, it is safe to say that children's weak phonemic representation for MSA words is predictable by virtue of their little experience with MSA and literacy.[26] However, one interesting question that has still gone unanswered is "Do CA words keep the advantage of their phonological representations over MSA words when diglossic speakers of Arabic minimize their linguistic distance and develop more stabilized knowledge of alphabetic literacy through MSA?" To answer this question, we need to know how adult literate speakers of Arabic react to the target phonemes in both varieties of Arabic.

4.3. Summary and Research Hypotheses

The present study examines the role of word frequency, neighborhood density, uniqueness point, cross-linguistic diglossia, and orthography in Arabic phoneme awareness. More specifically, this study will determine how fast and accurately literate speakers of Jordanian Arabic recognize target phonemes of the same linguistic affiliation (i.e. exists in both phonological inventories of Arabic) in the vocabulary of MSA and JCA. This work will test five proposed hypotheses.

The first hypothesis is built on the proposal that phoneme monitoring involves lexical access, and listeners commence phoneme detection after they recognize the carrier word. However, this does not entail that the listener has to wait until the end of the word to start phoneme searching. A more plausible proposal is that auditory lexical access is a gradual and proportional process of activation. That is, right after the UP the listener becomes fairly certain what the word is, e.g., 60%, and shifts part of his/her attention to phoneme identification, where he/she also obtains some certainty about whether the word contains the target phoneme, e.g., 60%. Towards the end of the word, the listener's

---

[26] The authors argue that equivalent phoneme transfer from children's CA phonological inventory is another reason for children's poor performance in MSA phoneme identification.

certainty about the word increases, e.g., up to 90%, especially if the word is frequent. Consequently, the listener becomes more confident about whether the word contains the target phoneme. Thus, the earlier the listener accesses the carrier word, the faster he or she will proceed to the phoneme detection stage. The current research hypothesizes a direct relationship between UPs and the time a listener takes to detect phonemes. Moreover, since both word frequency and duration are strong predictors of word recognition, this research postulates a direct relationship between word duration and response latencies to target phonemes, and an inverse relationship between word frequency and RTs to target phonemes. More specifically, this study hypothesizes:

> *H1.4: (a) the earlier the UP of a word is, the faster the listener would react to the target phoneme; (b) the shorter the spoken word is, the smaller RTs to the target phoneme would be; and (c) the more frequent the carrier word is, the faster the listener would detect the target phoneme.*

The second hypothesis predicts a stronger performance on phoneme monitoring when the target phoneme occurs in MSA words compared to JCA words. The present study introduces this hypothesis based on two previous findings. First, alphabetic literacy is a crucial factor for individuals’ development of phoneme awareness (Cheung et al., 2001; McBride-Chang et al., 2004; Morais et al., 1979). Second, orthographic representation of words is engaged in phoneme monitoring tasks (e.g., Dijkstra et al., 1995). Since MSA is the language of literacy, the one that Arabic speakers encode and decode with alphabetic orthography, we predict:

H2.4: adult diglossic speakers of Arabic detect the Arabic phonemes within MSA words faster than they detect the same phonemes within JCA words.

However, if the first hypothesis *(H1.4)* proves valid, we expect the literacy advantage for MSA to be reconciled or minimized by the processing advantage of JCA words found in Ibrahim and Aharon-Peretz's (2005) and duplicated in Chapter Three of this thesis. The rationale behind this prediction is that isolated CA words are processed faster than MSA words and lexical access precedes phoneme monitoring. So it would not be surprising to postulate:

H3.4: adult diglossic speakers of Arabic respond to a target phoneme in MSA words as fast as they respond to the same target phoneme in JCA words.

The fourth hypothesis pertains to the effects of phoneme identity on phoneme awareness. As stated earlier, beginning readers of Arabic are exposed to transparent phoneme-to-grapheme correspondences, but this shallow orthographic representation does not last for a long time. Usually, after the third grade, children start to use the less transparent (i.e. deep) orthography of Arabic where the diacritical marks that represent short vowels of Arabic are omitted from most Arabic texts. Since then, vowelized Arabic orthography is confined to children's stories and some religious scripts. As Arabic phonemes are different in their orthographic profiles, the present study predicts:

H4.4: speakers of Arabic detect Arabic consonants faster than Arabic short vowels.

The fifth research hypotheses will examine the prediction of LRM (Metsala & Walley, 1995; Walley et al., 2003) on the phoneme awareness of adult diglossic speakers of Arabic. According to LRM, the quality of phonological representation is word-specific, depending on listeners' familiarity and experience with the word as well as its phonological neighborhood density. In support, my proposed research hypothesizes that adult native speakers of Arabic will detect phonemes of high-frequency words faster than phonemes of low-frequency words when the other variables are held constant (the same as *H1(c)*, but built on a different justification). Moreover, adult native speakers of Arabic are expected to identify phonemes of Arabic words with many phonological neighbors easier than phonemes of Arabic words with few phonological neighbors when the other variables are partialled out. However, phonological neighborhood density can play an opposite inhibitory effect at the lexical access level as words of the same phonological family compete with the target word for selection (e.g., Goh et al., 2009; Luce, & Pisoni, 1989; 1998). If lexical access proves to be a prior step for phoneme identification, then it would be difficult to predict ahead of time which effect of neighborhood density would be stronger or whether they would cancel each other out. Thus, there are three possible contrasting sub-hypotheses:

H5.4 (a): adult native speakers of Arabic identify phonemes of Arabic words with many phonological neighbors faster than they identify phonemes of Arabic words with few phonological neighbors when the other variables are partialled out;

H5.4 (b): adult native speakers of Arabic are expected to identify phonemes of Arabic words with many phonological neighbors slower than they identify phonemes of Arabic words with few phonological neighbors when the other variables are partialled out;

H5.4 (c): adult native speakers of Arabic identify phonemes of Arabic words with many phonological neighbors as fast as they identify phonemes of Arabic words with few phonological neighbors when the other variables are partialled out

4.4. Methods

4.4.1. Participants

Thirty participants from the same population of the previous studies were recruited to participate in the fourth experiment for monetary compensation. The participants were native speakers of North Jordanian Arabic, a local form of rural Jordanian, and had their primary and secondary school education in Arabic. None of them had participated in the pretests or the other experiments. The participants reported normal hearing and reading abilities as a prerequisite for taking part in this experiment.

4.4.2. Materials and Design

The target phonemes consisted of two consonants and two short vowels. The consonant phonemes were the voiceless pharyngeal fricative /ħ/ and the voiced alveolar lateral /l/. The short vowels included the front high vowel /i/ and the back high vowel /u/. Each pair of consonants and short vowels is part of the MSA phonological inventory and exists in all vernaculars of Jordanian Arabic.[27] The carrier items were 144 JCA words (40 with /ħ/, 40 with /l/, 24 with /i/ and 40 with /u/) and 160 MSA words (40 with /ħ/, 40 with /l/, 40 with /i/ and 40 with /u/). The carrier words were composed of one to three syllables. Some words had phonological structures common to both varieties of Arabic (i.e., CV/CVC), other items had phonological structures that occur in JCA only (i.e., CCV/CCVC), and a third group of words included a syllable structure common in MSA but less common in JCA (i.e., CVCC; Al-Sughayer, 1990; Holes, 2004). The target phonemes varied in their positions in the carrier words. The researcher asked another group of the same population to rate the frequencies of the spoken carrier words on a 7-point scale using an online questionnaire.

---

[27] These sounds are just random examples of many other sounds present in both MSA and JCA.

Table 4.1: *Correlation matrix between the four continuous variables. Correlation Values larger than 0.4 were considered high and printed in bold.*

	Word Duration	UP	Rating Frequency	Neighbourhood Density
Word Duration	1.00	**0.55**	-0.08	-0.23
UP	**0.55**	1.00	0.04	0.26
Rating Frequency	-0.08	0.04	1.00	0.06
Neighbourhood Density	-0.23	0.26	0.06	1.00

Table 4.2: *Descriptive statistics for the tested variables.*

	Word Duration (ms)	UP (ms)	Word Frequency	Neighbourhood Density
MSA-Consonant				
Range	495-1365	279-854	2.5-6.32	0-18
Mean	745	542	4.54	4.7
S.D.	139	148	0.9	4
MSA-Short vowel				
Range	390-1000	229-768	2.5-5.3	0-28
Mean	672	449	4.2	4
S.D.	108	119	0.6	5
JCA-Consonant				
Range	444-1028	155-894	2.5-6.8	0-43
Mean	700	484	5	5.9
S.D.	141	133	0.9	7
JCA-Short vowel				
Range	333-831	226-745	3.5-6.3	0-32
Mean	635	417	5	5
S.D.	112	106	0.8	7.8

The researcher computed the phonological neighborhood densities for the MSA and JCA words by substituting, adding or deleting a single sound at any position in the word (Coltheart et al., 1977; Luce & Pisoni, 1998). A Python program generated all potential neighbors by changing one phoneme at a time. I manually looked through the list and counted all those I recognized as real words. The UPs for MSA and JCA words were identified based on *Al-Mawrid* (Baalbaki, 2008) and *muʕdʒam ʔalfaaðˤ ʔalħayaah lʕaamyah fi lʔurdun* (Dictionary of the Everyday Language in Jordan; 2006).[28] These dictionaries list the surface forms of MSA and JCA words, respectively, and provide the possible competitors for every word. After the UP phonemes had been identified, the carrier words were recorded. Then the UPs were measured in milliseconds, from the onset of the spoken word, both aurally and visually using a waveform editor. Following Radeau, Mousty, and Bertelson (1989), the beginning of the release noise was defined as the UP for oral stops, whereas the midpoint of the segment duration was defined as the UP for all other sounds. Table 4.1 demonstrates the correlation between the four continuous variables.[29] Table 4.2 summarizes the characteristic of the continuous variables across language and sound class conditions. The experiment included 304 filler items (i.e., words that do not contain the target phoneme; 144 JCA filler words and 160 MSA filler words). The filler words were comparable to the carrier words in the number of phonemes and syllable structures.

---

[28] I also checked the same dictionaries to assess the phonological neighborhood densities of the tested words.

[29] The correlation between word duration and UP is relatively high (0.5). However, this correlation is unproblematic, as the two variables were used to test the same hypothesis.

4.4.3. Procedure and Apparatus

A male native speaker of North Jordanian Arabic recorded the stimuli on a digital audio tape recorder. The subjects were asked to make a speeded response to the target phonemes occurring anywhere in the spoken words. The participants were instructed to press the designated button with their dominant hand if they heard the target sound, and to withhold the response if they did not hear the sound. Their reaction times (RTs) were collected with a SONY portable computer PC (CPU 2.40 GHZ) running Windows 7 and E-prime 2.0 presentation software (Psychological Software Tools, Inc., Pittsburgh, PA, USA; http://www.pstnet.com). The computer reported RTs, measured from the onset of the carrier word, and the response accuracy. The experiment consisted of two lists of items with two blocks in each list. List *A* comprised the voiceless fricative /ħ/ and the vowel /u/ while list *B* contained the voiced alveolar lateral /l/ and the vowel /i/. The participants were tested on only one experimental list and began the experiment with a block of 10 practice trials. The subjects were instructed to respond to the consonant target in the first block and to the short-vowel target in the second block of the experiment. There was a five-minute break between the two tested blocks. The participants were tested individually in a quiet room.

4.5. Results

Data from one participant were removed because he or she had overall error rates exceeding 30%. No items were rejected as a result of excessive error rates. Responses shorter than 500 ms and longer than 1500 ms, identified as upper and lower limits between outliers and the other responses, were also removed from the analysis (3% of the data).

To test the research hypotheses, I created mixed-effects models in R, version 3.2.3 with *lme4* (Bates et al., 2016) and *lmerTest* (Kuznetsova et al., 2016) packages. The mixed-effects models incorporated reaction time, measured in milliseconds, as the dependent variable, with three categorical and four continuous independent variables. The language of the carrier word (JCA versus MSA), phoneme type (consonant vs. short vowel), and syllable type (JCA-specific, MSA-specific, versus language-unspecific syllables) created a three-way categorical design. The model also included word duration, UP, word frequency, and neighborhood density (continuous variables), creating multiple regressions of correlational design. The participant, the carrier word, and the target sound were also incorporated into the model as random effects. I had the computer program include all variables at step one and eliminate the insignificant ones at each step until the final model was developed. Participants' error rates were also analyzed using a logistic function and binomial variance.

Figure 4.1 through 4.4 show the partial effect of a particular predictor, given that all other predictors are held constant at their means. For example, Figure 4.1 illustrates the relationship between word duration and RTs taking into account the hypothetical situation in which words are equally frequent, have the same UPs, and have the same neighborhood density, etc. Table 4.2 summarizes the results of the fitted model with the estimated coefficients in the second column, and their standard errors, *t*-values and *p*-values in the subsequent columns.

Table 4.3: *Summary of the mixed-effects model for both categorical and continuous variables predicting phoneme-monitoring time.*

Fixed Effect	Estimate	SE	*t*	*p(t)*
Intercept	940	25.4	37.0	< 0.0001
Language: JCA	41	20.3	2.0	0.044
Sound class: short vowel	119	5.8	20.5	< 0.0001
Word duration	0.25	0.03	8.9	< 0.0001
UP	0.04	0.03	1.24	1.241
Rating frequency	-36	3.4	-10.8	< 0.0001
Phonological density	0.8	0.9	0.88	0.380
Language: JCA* UP	0. 1	0.04	2.3	0.022
Language: JCA* Phonological density	-4	1.03	-3.8	0.0002

Figure 4.1 illustrates a direct relationship between RTs to the target phonemes and word duration. It shows that a word of 1000 ms delayed RTs to the target phonemes by approximately 250 ms. Figure 4.2 also depicts an adverse effect for UPs. When the UP occurred after 1000 ms from the onset of a JCA word, this slowed RTs to the target phoneme by approximately 130 ms. However, Figure 4.2 illustrates that the effect of UPs was weak for MSA words compared to JCA words. Inferential statistics, summarized in Table 4.3, demonstrates a non-significant effect for the UPs of MSA words ($p = 1.241$), and a significant interaction between UP and the language of the carrier word ($p = 0.02$), indicating a significant effect for UPs of the JCA carrier words. Word frequency had positive (i.e., facilitative) effects on the target phonemes regardless of the language of the

carrier words, as Figure 4.3 illustrates. The fitted model revealed that every point of increase in frequency rating significantly decreased RTs to the target phoneme by 36 ms ($p < 0.0001$). Generally, these results are consistent with *H1.4*, postulating significant effects for word length, UP, and word frequency on the speed of phoneme recognition.

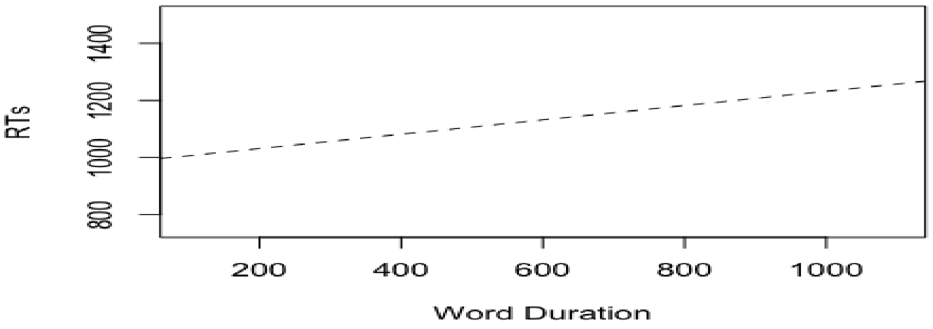


Figure 4.1: *Effects of word duration when RT is adjusted to the mean values of the other significant variables.*

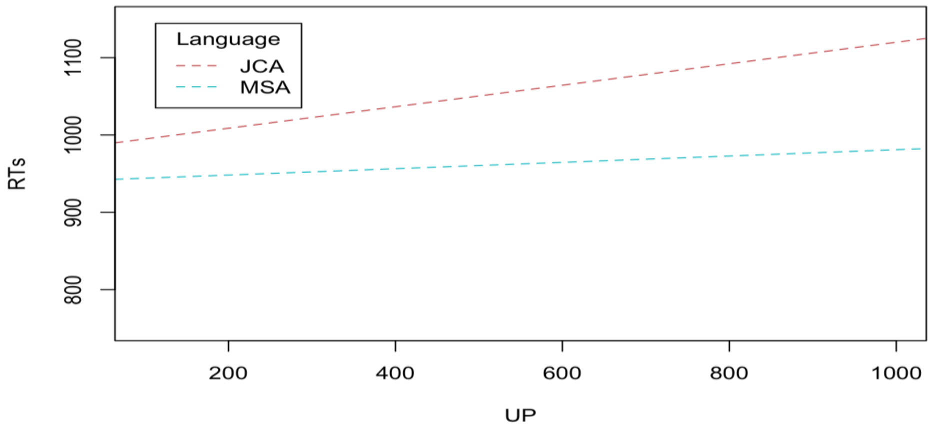


Figure 4.2: *Effects of UP across MSA and JCA words when RT is adjusted to the mean values of the other significant variables.*

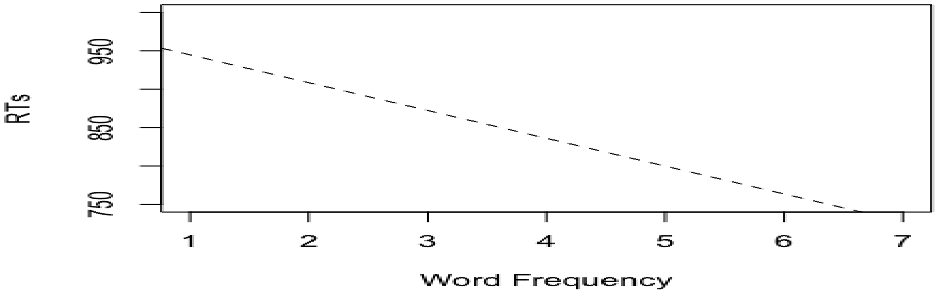


Figure 4.3: *Effects of word frequency when RT is adjusted to the mean values of the other significant variables.*

Figure 4.4 indicates a small difference in participants' RTs to the same target phonemes in MSA and JCA words. However, the figure shows that mean RTs to short vowels were noticeably larger than mean RTs to consonants. This considerable difference occurred in both MSA and JCA carrier words. Inferential statistics, listed in Table 4.2, validates the visual representation of the data. The table demonstrates that mean RTs to the target phonemes in JCA words were longer than mean RTs to the same target phonemes in MSA words by 41 ms. Nevertheless, the difference was marginally significant (p = 0.044) at $p \leq 0.05$ and non-significant at $p \leq 0.01$. This finding is more consistent with *H3.4*, which predicts a non-significant difference in RTs to the target phonemes occurring in JCA and MSA carrier words. The statistical analysis revealed a significant main effect for the sound class (p <0.0001) with faster responses to the target consonants relative to the short vowels (mean difference = 119 ms). The significant delay in reacting to a short vowel relative to consonants agrees with the fourth research hypothesis (*H4.4*). Results of the mixed-effects model, listed in Table 4.2, revealed a non-significant effect for the phonological neighborhood density of the MSA carrier words (p = 0.380), partially validating *H5.4(c)*, but a significant interaction between neighborhood density and language of the carrier word (p = 0.0002), partially confirming the first contrasting hypothesis (*H5.4(a)*). When the phonological family size of the JCA word increased by one phonologically-related word, this significantly corresponded to 3-ms faster responses to the target phoneme of that word. That is, a JCA word with 45 phonological neighbors could speed RTs by 135 ms (see Figure 4.5). This significant facilitative effect did not show up in the MSA carrier words. Finally, the syllable type variable was excluded from the final mixed-effects model. This is because RTs to the carrier words with language-specific syllables were not significantly different from RTs to the carrier words with syllables legitimate in both varieties of Arabic.

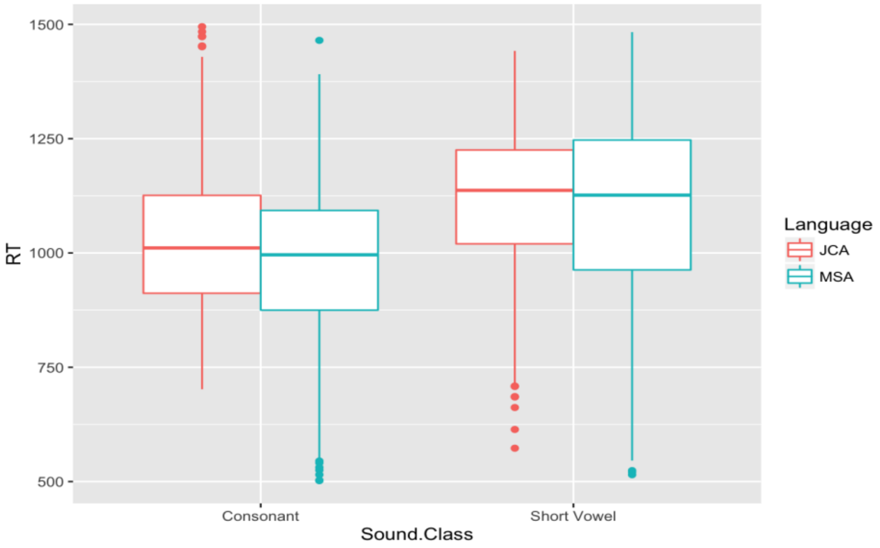


Figure 4.4: *RT by sound class and language of the carrier word.*

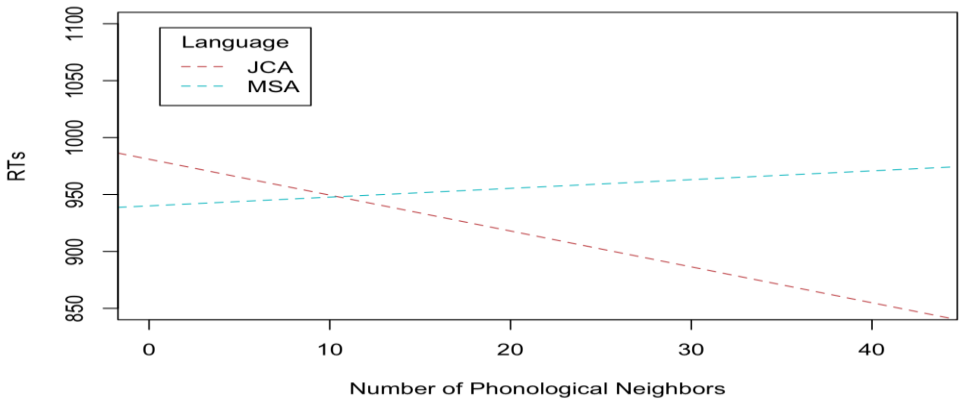


Figure 4.5: *Effects of phonological neighborhood density across MSA and JCA words when RT is adjusted to the mean values of the other significant variables.*

The analysis of accuracy performance shows very high accuracy rates for both language and sound class, as Figure 4.6 depicts. The participants recoded 97% accurate responses to the target consonants of MSA words and 94% accurate responses to the same consonant in JCA words. Of the total responses to the short vowels, there were 93% accurate responses to the target vowels in the MSA vocabulary and 86% accurate responses to the same target vowels in the JCA vocabulary.

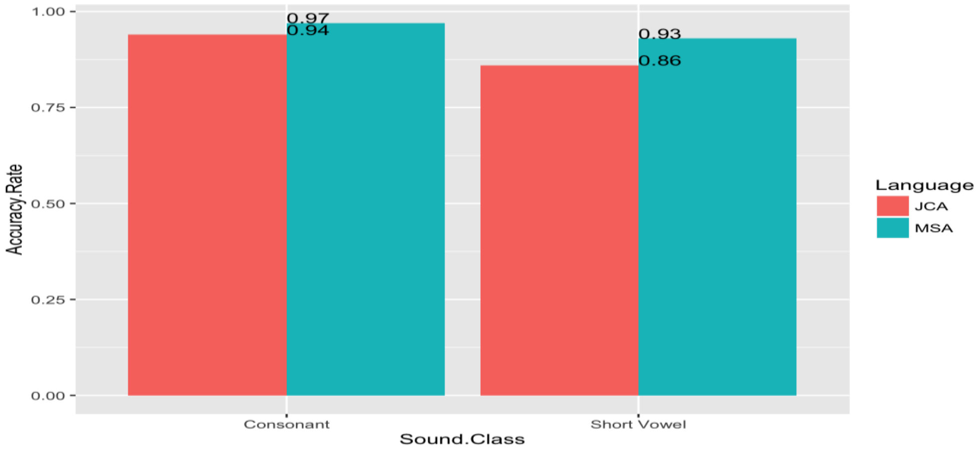


Figure 4.6: *Accuracy rates for sound class and language of the carrier word.*

Table 4.4: *Summary of logistic regression model for response accuracy, with correct responses scored as 1 and incorrect as 0. df = 8826*

	Estimate	SE	*z*	*p(z)*
Intercept	3.6807	0.1917	19.198	< 0.0001
Language: MSA	0.8064	0.1345	5.995	< 0.0001
Sound class: short vowel	-0.9446	0.2448	-3.859	0.00011

Table 4.3 sums up the logistic regression model used to analyze the small error data, with the estimated coefficients in the second column, and their associated standard errors, z- and *p*-values, in the subsequent columns. Results of the logistic model revealed significant differences between the two languages and sound classes. The logistic regression coefficients give the change in the log odds of the correct responses for MSA words compared to JCA words, and for vowels compared to consonants. In the fitted model above, the log odds of the correct responses for MSA were higher than the log odds for the JCA words by 0.8064. The vowel stimulus reduced the log odd by −0.9446, compared to the consonant stimulus. As stated earlier, these differences are small in the actual probabilities because the odds ratios for the accurate responses were already high (i.e., close to 100%) for both languages and phoneme types.

4.6. Discussion

The present study addressed the relationship between lexical information and phoneme awareness through an online phoneme monitoring task. To glean insight into this proposed relationship, the study investigated the role of three lexical factors in Arabic phoneme awareness: UP, word frequency, and word duration. Moreover, the present study addressed the relationship between the cross-linguistic situation of Arabic diglossia and the phonemic representations of Arabic words. This study raised the question of how lexical status (L1 vs. L2) and literacy affiliation (written vs. spoken) of the two varieties of Arabic can affect Arabic phoneme monitoring. The third purpose of this work was to examine the effect of Arabic orthographic depth in phoneme recognition.

This study has shown that phoneme monitoring goes through two successive stages of processing. In the first stage, listeners access and recognize the spoken word. Then they start to determine whether the recognized word contains the target phoneme. Some of the investigated factors contributed to the first stage of phoneme monitoring (i.e., the lexical stage), other factors involved the second stage of phoneme awareness (the post-lexical stage), and the third group of factors pertains to both stages. The Arabic phoneme monitoring task demonstrated inhibitory effects for both UP and word duration, whereas word frequency revealed a robust positive effect. Roles of these factors are very common in lexical decision tasks, as the three previous chapters and a large body of research reported (e.g., Cleland, Gaskell, Quinlan, & Tamminen, 2006; Goh, Suarez, Yap, & Hui Tan, 2009). Word frequency proved valid in both spoken word recognition (e.g., Cleland et al. 2006), and in offline phoneme awareness tasks such as phoneme manipulation (Hogan et al., 2011; Roth et al., 2006; Troia et al., 1996). The word frequency effect attested to in the present study can be attributed to its positive role in word recognition and its additional facilitative role in the post-lexical phoneme searching stage. At the lexical level, frequent words are easier to map onto their semantic representations. At the phonemic level, frequent words are more facilitative, as they have more segmental lexical representations (Metsala & Walley, 1995; Walley et al., 2003), perhaps because they vary in their pronunciation (Hogan et al., 2011). That is, since different speakers pronounce frequent words many times with different articulatory fine details, this variation may entrench the phonological representation of these frequent words. The adverse effects of UP and word duration and the robust positive effects of word frequency attested to in this experiment support the hypothesis that phoneme detection occurs post-lexically after having the target word processed at the lexical level. These results agree with the three models of word recognition introduced earlier in this study: the *TRACE model*, the *autonomous race model,* and the *Merge model.* All of these models predict that phonemes get (part of) their activation from the lexical node of the language apparatus.

However, the effects of UPs for MSA words found in the morphology experiment (Study I) and the priming experiment (Study III) were not replicated in the current experiments. Radeau, Morais, Mousty, and Bertelson (2000) observed for French that the effect of UP was restricted to slower speech rates. In the present experiment, it happened that the speaker recorded the JCA words and the MSA words in two different sittings. When the recoded stimuli were analyzed, it was found that the speaker's average speech rate for the MSA words was faster than his average speech rate for the JCA words by 2.4 syllables per second.[30] Thus, it is possible that this interaction is due to faster speech rates being characterized by more acoustic reduction.

More importantly, the present study found that participants reacted to the target phonemes in the MSA carrier words as fast as they responded to the same target phonemes in the JCA carrier words. This result should not be interpreted as if there were no advantage for MSA in phoneme processing. On the contrary, the findings confirm this

---

[30] The average speech rate for the MSA words in this study was 5.6 syllables per second.

advantage. Again, we can propose two stages of phoneme monitoring tasks to account for the observed processing balance. For each variety of Arabic, there is an advantage in one of the two stages and a disadvantage in the other stage. JCA is faster at the word recognition level. This is evident in Ibrahim and Aharon-Peretz's priming experiment (2005), which was replicated in Chapter Three. Alternatively, MSA is faster at the phoneme awareness level. That is, MSA has a stronger opportunity to scaffold the phonological representations of its words through their orthographic representations and through familiarity in print. The two influences pull each other in the opposing directions. MSA words delay word recognition but speed phoneme access, whereas JCA words speed word recognition as an L1 but slow phoneme tracking because they lack strong orthographic representations. This experiment suggests that the positive and negative effects of each stage neutralized the participant's performance in the vocabulary of both varieties of Arabic. These results do not agree with Russak and Saiegh-Haddad's (2011) finding that phoneme awareness is easier in the lexicon of L1 compared to the lexicon of L2.[31] The current analyses showed that the phonemic performance in L1 is not better than the phonemic performance in L2 when the L1 lacks salient orthographic representations. As for accuracy, literate native speakers of JCA scored very high accurate responses to the target phonemes in both languages, though their responses were somewhat more accurate to the target phonemes in MSA words (95%) relative to the same target phonemes in JCA words (90%). This indicates that literate speakers of JCA are not as poor at phoneme awareness of JCA words as illiterate adults or Hong Kong Chinese

---

[31] The authors investigated phoneme awareness in three different experiments: phoneme deletion, phoneme isolation, and phoneme segmentation. Since these are offline tasks, the lexical access stage (i.e., the first processing stage) attested to in the online phoneme-monitoring task becomes irrelevant.

children are (Cheung et al., 2001; McBride-Chang et al., 2004; Morais, 1991; Morais et al., 1979).

This study provided more evidence for the effects of Arabic literacy on phoneme awareness. It showed that literate native speakers of Arabic recognized Arabic consonants faster than Arabic short vowels. Arabic orthography is opaque because the short vowels correspond to diacritic marks rather than full letters, and these superscripted diacritics do not exist in most Arabic written material (e.g., handwriting, publications, printed media, and street signs). Previous studies emphasized that both Arabic and Hebrew native speaking children find it more difficult to isolate initial consonants in CV-units compared to the final consonants in VC-units (Saiegh, 2003; 2007a, 2007b; Saiegh et al., 2010). These studies ascribe the cohesiveness of the CV-units to the salience of the CV syllable in both Arabic and Hebrew (i.e., CV is by far the most frequent syllable in both languages). They also propose the visual representation of phonemes in the voweled Arabic/Hebrew orthography as another/alternative reason for the cohesiveness of CV-units. That is, superimposing the diacritics of short vowels onto the full letters of consonants (and sometimes long vowels) renders the CV phonological unit as one integral orthographic unit. Accordingly, this integrated orthographic unit reconstructs the consonant and the following short vowel as one holistic phonological unit, making them difficult to be segregated.

The present findings are consistent with the orthographic restructuring account of the CV integration. The results suggest that the non-salient and scarce orthographic representations for Arabic short vowels delay their recognition in the spoken modality relative to consonants. The findings show that effects of Arabic orthography on phonological representation appear not only in offline metalinguistic tasks (e.g., Saiegh-Haddad, 2004; 2007b) but also in online phoneme-monitoring tasks. Moreover, the results extend the evidence for the Arabic orthographic effects on phoneme awareness to incorporate adult perceptions. The present study demonstrated that the influence of Arabic deep orthography on phoneme monitoring is not confined to MSA words, which are commonly used in reading and writing. Short vowels also had a processing advantage over consonants in JCA bearer words. This indicates that the Arabic orthographic system reconstructs the phonological architecture of the Arabic lexicon in general, though the effect is stronger in MSA words. Pertaining to accuracy, it was somewhat surprising to find high accuracy rates for short vowels. Although native speakers of JCA were less accurate in detecting short vowels compared to consonants, their accuracy rates were still very high. This finding indicates large effects from small exposure. Arabs go to school and learn to read and write Arabic using diacritic marks for two or three years. Yet, this could be enough to give them accurate, but slow, phoneme awareness for the short vowels. This is similar to the performance of Chinese children in mainland China who can accurately delete and distinguish phonemes with only short-term exposure to Pinyin (i.e., alphabetic; Cheung et al., 2001; McBride-Chang et al., 2004). The high accurate responses to the short vowels of JCA words suggest that JCA short vowel awareness obtains free ride in MSA short vowel awareness. This claim is based on the fact that participants' early literacy education with diacritics has been exclusively in MSA and the diacritics disappeared long before they started to write in JCA (e.g., before they had their first cell phones).[32]

---

[32] None of the forty participants reported any ever experience with JCA words that were diacritically marked.

Phonological neighborhood density is another lexical factor attested to in this experiment. Word recognition and phoneme awareness are affected by how many phonologically similar words a speaker knows. Dense phonological family sizes have inhibitory effect at the word recognition level (e.g., Goh et al., 2009; Luce, & Pisoni, 1989; 1998) and a facilitative influence at the phoneme recognition level (Hogan et al., 2011; Metsala; 1999; Ventura et al., 2007). The adverse effect for dense phonological neighborhood results in a large number of similar words that would be activated and compete for selection when a listener hears the target word. Alternatively, words with large phonological families have stronger internal phonemic representations compared to words with sparse neighbors. This experiment revealed an interaction between phonological neighbors and the language of the carrier word. The experiment reported that neighborhood density facilitated phoneme awareness in JCA words but not in MSA words.

One possible account of the interaction between language and neighborhood density is that the facilitative effect of phonological neighborhood density at the phoneme recognition level outperforms its inhibitory effects at the word recognition level in the JCA words. The absence of a similar facilitative effect on the same target phonemes in MSA carrier words could be attributed to a possible balance in the two opposite effects. However, the nature of this observed interaction is a subject for further investigation. Another potential explanation for why the effect of neighborhood size is non-existent in MSA words is that the research estimates of the listeners’ neighborhoods may have been less useful for the MSA vocabulary compared to the JCA vocabulary. The researcher’s JCA vocabularies are probably fairly similar to the JCA vocabularies of the participants in this study. So the procedure for estimating of the neighborhood sizes using the JCA dictionary and the researcher's lexicon probably gave a reasonably accurate estimate of the average neighborhood sizes in the participants' lexicons. However, the vocabulary of MSA dictionaries could be substantially different (i.e., bigger) than the participants' MSA vocabulary, and participants' MSA words could even be very different from each other. This is possible if we take in to account that MSA is an L2, as a spoken language, for literate native speakers of Arabic. As a result, the procedure may have given a much noisier estimate of the neighborhood sizes for individual listeners, and we expect the relationship between the noisy estimate and response time to be weaker. Guessing the vocabulary of a typical experiment participant is still the trickiest issue in defining neighborhood size in languages such as English, and researchers have been thinking about it for decades.[33] In general, this finding agrees with the LRM proposal (Metsala & Walley, 1995; Walley et al., 2003). It partially confirms that words with dense phonological neighbors have strong segmental (phonemic) representations while words with sparse phonological neighbors have more holistic representations (Hogan et al., 2011).

A growing body of research has reported that literacy can affect how listeners process spoken language (e.g., Castles et al., 2003; Dijkstra et al., 1995; Hallé et al., 2000; Ranbom & Connine, 2011). However, there is still a debate about whether its effect is a result of online feedback or offline learning. According to the *bi-model interactive activation model*, orthography is automatically activated and feeds back information to the phonological systems of both lexical and sublexical levels (Grainger & Ferrand,

---

[33] The neighborhood density was also measured based on the consonantal roots only. Again, the root neighborhood density for the MSA words had no effects on RTs.

1996; Kiyonaga, Grainger, Midgley, & Holcomb, 2007). In the context of Arabic diglossia, it is possible that written forms of MSA words are co-activated more strongly than written forms of CA words. This is because written MSA is pervasive, whereas writing in CA is limited to electronic messages within the young population (Al-Khatib & Sabbah 2008; Haggan 2007; Mostari 2009). Similarly, this framework can explain how the phonological representation for a word with short vowels is formed through reading. Phoneme activation via spelling co-occurs with the generation of a phonological form consistent with the orthography. This framework proposes that phonological units are indirectly mapped onto orthographic units via an intermediate orthography-phonology interface at the prelexical level and directly at the lexical level. In the case of Arabic words, mapping short vowels onto orthography is weak or may not exist at all.

The other possible account of orthographic effect is based on the idea that learning about orthography can permanently change the way people perceives spoken language. Frith (1998) compared the acquisition of an alphabetic code to a virus that “infects all speech processing, as now whole word sounds are automatically broken up into sound constituents. Language is never the same again’’ (p. 1051). The rationale behind the emergence of orthographic effects on speech processing can be described within the framework of the lexical restructuring model (Metsala & Walley, 1998). In the context of Arabic diglossia, literacy reinforces the phonological representation of MSA words relative to CA words and consonants compared to short vowels. Learning to read in a language modifies existing phonological representations in the lexicon of that language by creating more detailed (i.e., finely specified) phonological representations for its words. The data suggest that reading in a language not only produces a more detailed phonological representation for its vocabulary but can also create an additional phonological representation that lacks features or segments present in the spoken form.

In sum, this study has shown that phoneme monitoring is a two-stage process. Listeners recognize the target word before they start to detect the target phoneme in an immediate follow-up stage. Sometimes, the two stages push our attention in the same positive direction, resulting in fast responses to the target phoneme. This is reflected in the profound effect of word frequency. The two stages may also resist each other in opposite directions. The balanced responses to the target phoneme in the vocabulary of both varieties of Arabic demonstrate the pros and cons of each variety in each stage. The literacy advantage of MSA words compensates their processing disadvantage at the lexical level. The effects of Arabic diglossia on phoneme awareness are not the same for school children and literate adults. While school children show a phoneme-awareness advantage for CA phonemes and words (Saiegh-Haddad, 2004; 2011), the present work revealed that with years of MSA literacy, MSA becomes more influential in phoneme awareness. More supporting evidence for the effect of literacy on Arabic phoneme awareness comes from the deep orthography of Arabic. The dominant unvoweled orthography of Arabic explains the significant delay in detecting Arabic short vowels compared to consonants.

Chapter Five

General Discussion and Conclusions

5.1 Introduction

This research was designed to investigate the effects of various linguistic components on Arabic lexical access in the aural modality. It revealed that morphological, sentential, semantic, phonological, and orthographic constituents of the Arabic language could influence Arabic language processing. The present work shows some effects of Arabic morphology and diglossia on word recognition. This dissertation consisted of four separate studies from perspectives of four different metalinguistic contexts. This chapter summarizes the research results and discusses their implications for theories of lexical representation and processing. The results obtained have implications for theories of (Arabic) morphological processing and representation, theories of bilingual lexical activation, theories of Arabic diglossia, and theories of orthographic and literacy knowledge on spoken word/phoneme recognition. This final chapter also proposes some future experiments to help establish the correct interpretation of some of the results presented in this thesis. The final section of this chapter concludes with how this project contributes to the broader scholarship of language processing.

5.2 Summary of Results and Implications for Theories of Lexical Processing

A number of previous psycholinguistic studies on Arabic morphology found priming effects between words that share the same word patterns and, more importantly, words that share the same roots. Some research concluded that effects of root priming were independent of the degree of semantic overlap between the prime and the target. Researchers introduced these results as evidence for the nonconcatenative view of Arabic morphology. That is, roots and word patterns are abstract morphemes of Arabic. The first experiment of this thesis focused on the role of Arabic morphology in lexical access. It investigated the role of Arabic morphology in lexical processing directly through a single lexical decision task rather than through priming techniques. This method paves the way to examine the role of two possible approaches of Arabic morphologies; namely, the root-based approach (Bauer, 2004; Cantineau, 1950; McCarthy, 1981) and the stem-based approach (Benmamoun, 1999, 2003; Ratcliffe, 1998, 2004). A group of literate native speakers of Arabic reacted to a number of MSA words varied in their whole-word frequencies, cumulative root frequencies, cumulative stem frequencies, root family sizes, and stem family sizes. The analysis revealed that the statistical model that incorporated all of the three types of frequencies and the root family size accounted for more explanatory data than any other complex or simple model. The higher the values of these variables in a word, the faster the response times to that word would be. The data showed that words, roots, and stems all played comparable roles in Arabic lexical access through their frequencies. According to these results, native speakers of Arabic activate or access the Arabic word *taħadʒdʒar* ‘to petrify’ as one whole unit, as a prefix-stem constituent [ta-ħadʒdʒar], and as a root-pattern constituent {ħdʒr/cvcaccac}. All of this suggests that Arabic is both a root-based and a stem-based language.

As discussed in Chapter One, these results can also be used to evaluate four major theories about how lexical items are processed and represented. The first theory is the whole word hypothesis. Proponents of this hypothesis suggest that words are fully listed in our lexicon (i.e., mental dictionary), and words are the only basic units of lexical representation. The second theory is the morpheme-based theory. This theory is in sharp contrast to the full-listing hypothesis. It proposes that words are accessed in the central system only through their obligatory decomposed morphemes. The third theory is the dual access hypothesis (e.g., Burani & Caramazza, 1987; Schreuder & Baayen, 1997). This account suggests hybrid processing in which words are accessed directly as whole units and indirectly through their decomposed morphemes. The fourth main theory for lexical processing is the distributed connectionist theory (Gonnerman, 2000). According to this theory, words are connected through networks and the network for a set of words becomes stronger when their shared components are more frequent and more consistent. (Davis et al., 2003; Gonnerman et al., 2007). The results are compatible with the last two hypotheses (i.e., the dual route hypothesis and the distributed connectionist hypothesis), but not with the first two hypotheses (the word-based hypothesis and the morpheme-based hypothesis). This is because the whole-word frequency and the morpheme frequencies together accounted for more data (participants' latency times) compared to any single account of frequency.

Turning to the issues of Arabic diglossia, the second study was designed to determine whether the language-processing bias that pertains to classical bilingualism could also be relevant in the context of Arabic diglossia. Diglossic speakers of Arabic use two coexisting varieties of the same language rather than typical independent languages. Previous studies on bilingualism found that unbalanced bilinguals process their first language (L1) faster than their second language (L2). Other works found that bilinguals recognize target words in language-consistent contexts faster than the same target words in language-switching contexts (Grosjean, 2008; Soares & Grosjean, 1984). Native speakers of Arabic use two varieties of Arabic, colloquial Arabic (CA) and Modern

Standard Arabic (MSA), in different contexts of their daily life. Although each of these two varieties has its domain of use, literate speakers of Arabic sometimes alternate between them at the word level to serve different functions. This study addressed the question whether native speakers of Arabic experience a lexical switching cost when they process an Arabic target word hosted in a carrier sentence of the other variety of Arabic. It also asked whether MSA is an L2 for literate native speakers of Arabic, as Ibrahim and Aharon-Peretz (2005) proposed.

To test the research hypotheses, a group of literate native speakers of Jordanian Arabic performed a lexical decision task in sentential contexts (i.e., to decide whether the last word of a sentence has meaning or not). The carrier sentences ended with a target word of either the same or the other variety of Arabic. The data showed that the participants reacted significantly faster to the target words in the non-switching conditions compared to the same target words in the language switching conditions. This finding is consistent with the BIMOLA model of word recognition (Grosjean, 1997; 2008). This model suggests possible feedback from the higher language level of the processing hierarchy to the lower lexical and sublexical levels. This switching cost indicates that the base language (i.e. the carrier sentence) pushes the participant to think of code switches as base language words.

Contrary to expectations, the participants reacted 24 ms faster to the MSA words than they reacted to the CA words. This latter finding is of particular interest because it makes it impossible to expect language processing just based on the simple view that L2 has to be disadvantaged. It is possible that the differences in the literacy status of the two lexicons of Arabic were responsible for the processing advantage of the MSA vocabulary. In the context of Arabic diglossia, MSA has a stronger orthographic representation compared to JCA because MSA is the language of literacy for speakers of Arabic. This could make literate native speakers of Arabic take less time to separate the final word out of the carrier sentence, as a metalinguistic judgment, when it is a MSA word. Accordingly, it is proposed that literacy facilitates some metalinguistic tasks such as word segmentation (i.e., identifying spoken word boundaries in sentential contexts), which occurs prior to lexical access in the context of this experiment. As the second experiment was indecisive about whether MSA is an L2 for literate native speakers of Arabic, the third study was devoted to answering this initial question through replicating Ibrahim and Aharon-Peretz's experiment on a group of literate native speakers of Jordanian Arabic.

The third study compared semantic priming effects within and across MSA and JCA words. The results showed that JCA target words were accessed faster than MSA targets when the primes were unrelated to the targets. Moreover, effects of semantic priming were greater when the primes were presented in JCA relative to effects of semantic priming when the primes were presented in MSA. The priming effect within JCA pairs was also larger than the priming effect between languages. These results are consistent with Ibrahim and Aharon-Peretz's findings (2005) on Arabic-Hebrew bilinguals of Israel. The reproduced results challenge Boudelaa and Marslen-Wilson's (2013) proposal that MSA is an L2 for Israeli Arabic-Hebrew bilinguals because of their insufficient exposure to MSA compared to the other Arabic speakers living in the rest of the Arabic-speaking world. The findings could be explained through the *revised hierarchical model* (RHM, Kroll & Groot, 1997; Kroll & Stewart 1994), which predicts stronger links between L1 words and their semantic representations compared to their L2 translation equivalents.

The fourth study compared the phoneme monitoring of consonants and short vowels in MSA and JCA carrier words. An important aspect of Arabic diglossia is that MSA words have strong orthographic representations, as it is the language of literacy, whereas any CA variety has a very marginal role in reading and writing. Arabic consonants have salient orthographic representations while the orthographic representations for the short vowels are diacritical and sporadic. The results confirmed the hypothesis that phoneme monitoring is a word recognition process in which the listeners access the lexical representation of the word before they decide on the phonemic representation of that word (e.g., Dijkstra et al., 1995; Norris et al., 2000). A large body of empirical research reported significant effects for word frequency, UP, and word duration on lexical access. The present data showed a similar effect for these factors in the phoneme monitoring task. Results of this study also agree with the general assumption that sounds and words that correspond to alphabetic representations have stronger phonemic representations relative to sounds and words that have no or occasional alphabetic forms (e.g., Cheung et al., 2001; McBride-Chang et al., 2004; Saiegh-Haddad, 2007). The experiment demonstrated that literate native speakers of Arabic detect Arabic consonants, which correspond to salient and frequent orthographic representations, faster than Arabic short vowels, which have weak and occasional representations. Most importantly, the fourth study reported no significant differences in participants' RTs to the target phoneme in the two varieties of Arabic. This latter finding is of particular interest, as it concurs with the two aforementioned general assumptions.

At the lexical access stage, JCA words are accessed faster than MSA words are (Study III). However, because languages with alphabetic orthographies have stronger phoneme representations compared to languages without (alphabetic) orthographies, then phonemes of MSA words are identified faster than phonemes of JCA words at the phoneme awareness level. Consequently, it is possible that JCA outperforms MSA at the lexical access stage, whereas MSA outperforms JCA at the phoneme identification stage. As a result, the two opposing advantages cancel each other.

Results of the three experiments on Arabic diglossia together underscore two general conclusions. First, MSA is an L2 for literate native speakers of Arabic. In general, literate native speakers of Arabic access the spoken vocabulary of CA faster than they access the spoken vocabulary of MSA (see Study III). This finding is based on the view that there is earlier and more frequent exposure to CA, compared to MSA, as oral means of communication. Second, it is impossible to expect language processing just based on the simple view that L2 has to be disadvantaged. This research concluded that some metalinguistic skills rely on the literacy status of the target words over and above their spoken familiarities. This includes the ability to mark word boundaries in spoken carrier sentences (Study II) and the capacity to identify the target phoneme in spoken carrier words (Study IV). In such auditory metalinguistic tasks, the advantage of MSA as the dominant language of literacy competes with and sometimes overcomes the advantage of CA as the dominant language of speaking.

Outcomes of the three experiments suggest that these tasks are processed differentially. The priming task seems to be motivated by the listening and speaking competence, resulting in a processing advantage for the spoken language over the literary one. However, recognizing target-word boundaries in spoken sentential contexts seems to be elaborated via the literacy competence. This is because the task involves separating a target word from the rest of the utterance. So it appears that identifying word boundaries is easier if spoken words correspond to written codes in the listener's visual experience. Phoneme monitoring tasks tend to be motivated via the spoken language experience at the lexical access level and via the literary language experience at the phoneme identification level. As each variety of Arabic (i.e. spoken vs. literary) has a processing advantage over the other in only one of the two necessary stages in a phoneme monitoring task, the two opposing advantages cancel each other out. Consequently, listeners identify the target phoneme in MSA words as fast as they identify the same target phoneme in JCA words. This research emphasizes that literacy and spoken language experience both contribute to the phonological and lexical awareness in perceiving speech. The data may suggest two possible loci for the literacy/orthographic effects in spoken word recognition. It is possible that the orthographic codes are activated during word recognition, or that the literacy introduces new representations of language. According to the latter assumption, phonological and holistic representations of words are affected by orthography during literacy acquisition.

5.3 Future Research

The suggested implications for future research are based on what this work has not done, as well as the questions that it has prompted. Below is a short list of directions for future research.

a. The results of Study I show positive effects for both stems and roots in Arabic lexical processing. These findings suggest that Arabic has two coexistent types of morphologies:

concatenative and nonconcatenative. However, it is possible that one of these morphologies has been externally superimposed on the lexical representation of native speakers of Arabic. Perhaps Arabic is a stem-based language, and the cognitive reality of roots is the outcome of school instructions. Alternatively, it could be that Arabic is a root-based language, and the cognitive reality of stems is the result of participants' good command of English. To reach a more decisive conclusion, future research may reexamine effects of Arabic roots and stems on monolingual illiterate speakers of Arabic, who do not know literary Arabic or any L2 at all, using words from their dialectical variety of Arabic.

b. The results of Study II indicate a chance of potential interaction between the language of the target word and the language of the carrier sentence, but the experiment might not have enough power to detect the interaction. This study can be replicated with a larger sample of stimuli. If a larger sample of data shows the proposed interaction, then *H3.2* will be confirmed: the magnitude of switching cost from MSA into JCA is significantly larger than the magnitude of switching cost from JCA into MSA. This possible hypothesis is based on the belief that switching into JCA is less common than switching into MSA.

c. Study IV argues that MSA words have stronger representations at the phoneme-level, as opposed to the lexical-level, than CA words do. However, this advantage did not appear on the surface because the lexical advantage of the JCA words and the phonological advantage of the MSA words cancel each other out in online phoneme monitoring tasks. Further research should compare the phonological representations of MSA words to the phonological representations of CA words using offline phoneme awareness tasks, where lexical processing (i.e., step one) becomes irrelevant. For example, future work may ask how well literate speakers of Arabic can manipulate (delete or segment) target phonemes in literary (i.e., MSA) and spoken (i.e., CA) words. On the basis of their literary experience, literate speakers of Arabic are predicted to manipulate the target phoneme in MSA words more easily than the same target phoneme in JCA words.

5.4 Contributions

This project provides some new insights into the way diglossic speakers of Arabic process their language. It concludes that Arabic speech processing is a complex mechanism and the result of multiple sources of information: information from the surface structure of words, information from their decomposed morphemes, information about how early they are acquired and experienced, and information about their literacy status. This research demonstrates that Arabic speech processing involves both bottom-up and top-down mechanisms. This section will summarize the bigger picture aspects of Arabic speech processing and how the results fit in the field broadly.

This thesis suggests that Arabic diglossia is a case of bilingualism from a psycholinguistic point of view. It demonstrates some analogies between Arabic diglossia and previous findings on typical bilingualism in auditory word recognition. Diglossic speakers of Arabic are biased toward processing Arabic target words in the context of the same variety of Arabic more easily than processing Arabic target words in the context of the other variety of Arabic. This finding indicates that when native speakers of Arabic access a target word, they access it based on the lexicon of the base language. When the target word is a code-switched word, the listeners switch their search to the other lexicon of Arabic. However, what remains unclear is the locus of switching cost in the context of Arabic diglossia. One possibility is that language processing is part of the general cognitive controls (Garbin et al., 2010; Abutalebi et al., 2012). Alternatively, it is possible that diglossic speakers of Arabic, and typical bilinguals, develop language-specific mechanisms to control language (Abutalebi et al., 2008; Calabria et al., 2011). In a recent work, Blanco-Elorrieta and Pylkkanen (2016) used magnetoencephalography (MEG) to monitor the brain activity of proficient Arabic-English bilinguals while they were processing language switching vs. category switching in both perception and production. The neuroimaging data have shown that language switching in comprehension recruits the anterior cingulate cortex while language switching in production and the categorical (i.e., non-linguistic) switching in both comprehension and production activate the dorsolateral prefrontal cortex. This difference suggests that the general cognition control is responsible for language switching in production, whereas language switching in comprehension is dominated by a language-specific control mechanism. Similar neuroimaging research is needed to determine whether the language switching cost found in the switching task of Arabic diglossia (Study II) is motivated by the general cognitive control or a language-specific control.

The present research has also revealed an analogy between how early bilinguals access their L2s and how diglossic speakers of Arabic access their MSA lexical items. The results demonstrate a MSA-CA priming effect, though weaker than CA-MSA priming. A similar pattern of priming has been found between French primes (L2) and English targets (L1) in both simultaneous and early English-French bilinguals but not in late English-French bilinguals (Sabourin et al., 2014). Accordingly, results of the Arabic diglossia experiment indicate that MSA lexical items are integrated into the CA-based semantic network because diglossic speakers of Arabic acquire the two varieties in early succession. However, Sabourin and colleagues' results are more decisive because their experiment was a visual masked-priming task, a more (subconscious) task, while Study II employed a regular auditory priming experiment. To obtain more valid results about how effective MSA senses of translation equivalents incorporate into the L1-based semantic network, future research should eliminate the task strategic effects. A future version of this experiment can employ more subconscious auditory priming techniques where the volume of the prime is attenuated and compressed (Ussushkin, 2015).[34]

This work contributes to the body of research investigated the effects of literacy on speech processing. Previous findings have shown that the lack of (alphabet) literacy for pre-readers and illiterate adults adversely affects some aspects of speech processing such as phoneme awareness and word segmentation. The current thesis further revealed that literate adults who receive their literacy in one language but not in the other develop a stronger phonological representation for the language of literacy even if it is their L2. This finding presents some new evidence for the interface between spoken- and written-language experience in speech processing.

Part of this research contributes to the body of literature on Arabic morphology by exploring how Arabic morphology affects word perception from both non-linear and linear standpoints. More specifically, the research here contributes to the literature on Arabic morphology processing in two ways. First, the current research supports previous findings and claims from morphological priming tasks regarding the positive effect of Arabic consonantal roots in word recognition. Second, the current findings give an insight into the relevance of stems and whole words in Arabic word recognition. What remains to be seen is why diglossic speakers of Arabic develop two distinct paths of

---

[34] Many thanks to Dr. Laura Sabourin for drawing my attention to this masked priming technique in auditory word recognition.

morphology. If it turns out that Arabic is a stem-based language and the Arabic root is the byproduct of literacy instructions (i.e., formal teaching of language structure), then the role of Arabic literacy in speech processing would not be limited to phoneme awareness and word boundary identification; but it would further extend into lexical access. Be that as it may, this research proposes that the dual model of word recognition can be dual not only in the sense of having surface and morphologically decomposed representations of the word, but also in the sense of its ability to incorporate two distinct paths of morphology.

Appendix A
Arabic Phoneme Inventory

Consonants

Arabic Consonants	IPA Symbol	Description
ء	ʔ	Glottal stop
ب	b	Voiced bilabial stop
ت	t	Voiceless dento-alveolar stop
ث	θ	Voiceless inter-dental fricative
ج	dʒ	Voiced post-alveolar affricate
ح	ħ	Voiceless pharyngeal fricative
خ	x	Voiceless velar fricative
د	d	Voiced dento-alveolar stop
ذ	ð	Voiced inter-dental fricative
ر	r	Voiced alveolar tap
ز	z	Voiced alveolar fricative
س	s	Voiceless alveolar fricative
ش	ʃ	Voiceless alveo-palatal fricative
ص	sˤ	Voiceless alveolar emphatic fricative
ض	dˤ	Voiced alveolar emphatic stop
ط	tˤ	Voiceless dento-alveolar emphatic stop
ظ	ðˤ	Voiced inter-dental emphatic fricative
ع	ʕ	Voiced pharyngeal fricative
غ	ɣ	Voiced velar fricative
ف	f	Voiceless labio-dental fricative
ق	q	Voiced uvular stop
ك	k	Voiceless velar stop
ل	l	Voiced alveolar lateral
م	m	Voiced bilabial nasal
ن	n	Voiced alveolar nasal
ه	h	Voiceless glottal fricative
و	w	Voiced labio-velar glide
ي	y	Voiced palatal glide
“colloquial”	g	Voiced velar stop

Vowels

Arabic Vowel	IPA Symbol	Description
◌َ	a	Central, nearly half-open, unrounded, short
◌ُ	u	Back, nearly close, rounded, short
◌ِ	i	Front, nearly close, high, unrounded, short
آ	aa	Central, nearly half-open, unrounded, long
و	uu	Back, nearly close, rounded, long
ي	ii	Front, nearly close, high, unrounded, long
“colloquial”	e	Front, half close, unrounded
“colloquial”	o	Back, half open, rounded

Appendix B
Items of Study I with Their Characteristics

Word	Surface Frequency	Root Family Size	Root Cumulative Frequency	Root Family Neighbors More Frequent than the Stimulous	Stem Family Size	Stem Cumulative Frequency	Stem Family Neighbors More Frequent than the Stimulous	UP	Word Duration
مِرْوَد [mirwad] eye pencil	0.03	34	372.33	22	1	0.07	0	674	571
كاحِل [kaaħil] ankle	0.08	22	10.29	2	1	0.46	0	515	453
تَخْمِير [taxmiir] fermentation	0.05	21	18.62	8	1	0.16	0	685	623
تَقرِيع [taqriiʕ] scolding	0.18	25	53.68	4	1	1.22	0	634	543
حُرْقَة [ħurqah] agony	0.31	24	168.23	8	1	1.48	0	657	418
نُزُوع [nuzuuʕ] inclination	1.59	24	397.46	5	1	8.12	0	577	518
لَزوم [luzuum] need	4.63	29	383.3	7	1	8.91	0	720	611
فَخُور [faxuur] proud	1.35	27	112.5	4	1	4.18	0	727	431
بَلاء [balaaʔ] adversity	1.64	26	81.51	2	1	5.26	0	699	589
خَضْراء	7.18	21	119.89	1	1	25.48	0	873	747

[xadˤraaʔ] green.Feminine									
جَلِيس [dʒaliis] associate	0.13	12	2535.35	4	1	0.99	0	942	716
عَوْم [ʕawm] floating	0.16	11	4439.27	3	1	0.79	0	539	466
تَتِمَّة [tatimmah] continuation	1.01	14	1595.88	4	2	3.98	0	642	233
مُصْحَف [musˤħaf] The Holy Koran	0.78	12	1094.71	3	1	4.9	0	757	302
ثُمامَة [θumaamah] millet	0.03	4	1136.41	0	2	0.16	0	822	624
مَتْجَر [matdʒar] shop	3.36	12	1157.11	2	1	9.5	0	609	315
عَلانِيَة [ʕalaaniyah] publicity	3.23	12	1575.18	7	2	6.01	0	750	428
مواظِب [muwaaðˤib] diligent	0.05	4	3.32	0	4	3.36	2	841	489
شاكَسَ [ʃaakasa] to quarrel with	0.03	9	4.8	3	7	4.87	2	849	576
[واءَمَ] waaʔama to suit	0.03	7	21.68	3	6	9.71	6	645	444
جازَفَ [dʒaazafa] to take a risk	0.18	5	9.99	0	4	10.09	1	740	513
تَلاطَمَ [talaatˤama] to clash each other	0.13	8	6.67	2	5	2.04	1	796	433
مُمازَحَة [mumaazaħa] joking	0.13	7	10.12	2	6	2.53	1	874	491
أَجَّجَ [ʔadʒdʒadʒa] to inflame	0.19	10	10.81	2	6	4.14	2	651	425

مُفَرقَعات [mufarqaʕaat] fireworks	0.36	4	3.39	0	4	3.42	0	1188	461
مُثَلّج [muθalladʒ] frozen	0.1	11	26.82	2	5	1.77	0	767	663
مُفاجأة [mufaadʒaʔah] surprise	20.21	11	139.65	1	7	86.04	0	841	530
تَوَرَّطَ [tawarratˤa] to get involved	15.53	10	84.31	1	6	75.21	1	766	569
ناشَدَ [naaʃada] to entreat	6.55	12	109.08	0	5	52.08	0	706	460
سَطّرَ [satˤtˤara] to draw lines	2.05	11	40.04	2	4	4.04	0	708	531
مُهَيَّأ [muhayyaʔ] prepared	2.31	11	38.2	1	6	30.86	1	846	697
تبادَرَ [tabaadara] to cross someone’s mind	4.76	16	322.28	1	5	257.79	2	806	620
تَوْزِيعة [tawziiʕah] a delivery	0.03	9	293.39	0	2	183.6	1	830	459
أسلوبيَّة [ʔusluubiyyah] stylistics	1.02	19	584.01	0	3	224.4	1	966	390
تاسِع [taasiʕ] nineth	0.47	11	321.32	0	1	104.9	0	632	307
شَوط [ʃawtˤ] round	2.78	2	80.34	0	1	73.81	0	607	506
جَوهَرَة [dʒawharah] diamond	1.56	6	207.52	0	3	180.6	1	798	462
خَشَبيّ [xaʃabiyy] wooden	1.25	10	69.34	1	3	64.01	1	768	377
حَماسَة [ħamaasah]	6.81	10	243.23	0	3	207.0	1	839	619

Enthusiasm									
تَدريب [tadriib] training	35.81	12	467.61	0	2	274.0	0	881	739
خُطوة [xutˤwah] step	94.36	7	423.11	0	1	383.1	0	711	558
صَعْب [sˤaʕb] hard	24.06	11	405.56	0	1	227.6	0	753	595
هرَجَان [mihradʒaan] festival	52.93	2	175.65	0	1	177.1	0	890	403
يَنبَغي [yambaɣiy] ought to	108.82	14	236.46	0	1	154.1	0	751	479
رَقَم [raqam] number	89.68	10	341.45	0	2	232.9	0	469	386
كِيان [kiyaan] entity	21.59	6	117.15	0	3	112.5	0	699	599
قَطّاعَة [qatˤtˤaaʕah] a pair of pliers	0.03	53	1838.65	14	2	3.95	1	756	593
مُقتَبَل [muqtabal] prime of life	1.69	50	5073.8	20	1	2.17	0	805	711
مِقْدَام [miqdaam] courageous	0.18	43	4173.42	18	1	0.85	0	727	644
بَائِن [baaʔin] obvious	0.1	43	4786.73	19	2	0.85	0	643	560
ظِهَار [ðˤihaar]	0.05	38	1824.82	20	1	0.1	0	719	719
مَعْمَل [maʕmal] workshop	10.12	37	8294.17	8	2	26.83	0	662	577
عِداد [ʕidaad] among	10.69	36	5182.82	9	1	16.21	0	784	678
فَهيّم [fahiim] intelligent	10.33	23	832.09	4	1	13.71	0	770	395

تَسيير [tasiyyr] handling	12.2	25	1457.73	6	1	27.62	0	901	819
طائِلة [tˤaaʔilah] great	10.51	28	956.08	4	2	20.25	0	597	402
مُلَبَّس [mulabbas] covered with	0.36	29	167.54	4	6	8.81	1	974	793
مُبَطَّن [mubatˤtˤan] lined	0.49	25	41.57	4	4	3.32	0	803	385
تَحَجَّرَ [taħadʒdʒara] to petrify	0.34	21	231.61	6	5	7.24	2	700	629
مُعَجَّل [muʕadʒdʒal] accelerated	0.55	22	161	6	6	9.99	1	808	711
قامَرَ [qaamara] to gamble	0.08	21	114.06	3	6	3.35	2	566	426
تَصَلّب [tasˤallub] hardening	2	24	131.12	3	5	9.99	0	814	707
مُتَعَصِّب [mutaʕasˤsˤib] fanatic	1.09	24	196.16	4	8	32.7	1	866	810
تَغَيَّبَ [taɣayyaba] to absent himself	9.29	28	340.98	2	6	23.21	0	718	556
تَعاطُف [taʕaatˤuf] sympathy	5.18	32	257.48	1	5	42.25	0	827	653
عِمادَة [ʕimaadah] deanery	0.39	33	414.61	0	3	74.76	2	824	684
مَمات [mamaat] death	0.1	24	346.81	4	1	51.16	1	731	612
رَاوِي [raawiy] narrator	0.42	24	420.6	2	1	24.69	0	577	460
مُبْرَم [mubram] irrevocable	0.7	23	81.25	0	3	24.69	0	700	578

نِقاب [niqaab] veil	0.13	23	378.15	0	3	269.4	2	667	667
دَقِيق [daqiiq] fine	17.43	28	464.48	0	3	219.2	0	727	727
بَسِيط [basiitˤ] simple, humble	16.85	25	275.84	0	2	72.09	0	690	599
زَهرَة [zahrah] rose	7.33	25	172.56	2	3	21.92	0	471	434
قُرُوض [quruudˤ] loans	24.92	25	226.79	0	1	129.9	0	666	540
خَيال [xayaal] imagination	9.75	24	198.01	0	2	65.03	0	864	775
مُنَقَّط [munaqqatˤ] dotted	0.08	9	738.01	0	4	0.79	0	757	656
عانَدَ [ʕaanada] to oppose	0.13	14	1549.4	2	7	2.99	0	676	486
أزَّمَ [ʔazzama] to aggravate	0.05	8	1115.38	2	4	15.45	2	600	412
سَلَّطَ [sallatˤa] to inflict	0.91	10	1117.3	2	6	32.23	0	640	489
عانَقَ [ʕaanaqa] to hug	0.34	12	27.96	3	7	8.22	0	641	450
حارَبَ [ħaaraba] to fight	2.68	14	1324.66	4	7	79.2	1	667	493
مُثَقَّف [muθaqqaf] educated	5.72	11	566.86	1	6	147.9	0	852	712
تَكَيَّفَ [takayyafa] to adjust himself to	1.35	12	542.38	3	6	27.56	1	747	562
أهرَام [ʔahraam] pyramids	2.47	7	1189.24	0	1	280.4	0	821	498

قاهِر [qaahir] conqueror	0.73	10	770.82	0	1	726.4	0	567	269
حَرْبَة [ħarbah] bayonet	3.1	14	1324.66	0	3	1042	0	520	431
مَساس [masaas] violation	3.54	13	1975.14	3	1	25.45	0	891	683
ماضٍ [maadˤin] past	5.28	11	1629.11	1	2	1482	1	677	551
لِجان [lidʒaan] committees	37.09	5	1050.31	0	1	1472.9	0	774	649
هَدَف [hadaf] target, goal	77.56	12	1249.79	0	1	599.2	0	655	472
عَنَاوِين ʕanaawiin headlines	10.04	4	1586.7	0	1	31.96	0	853	442
تَطْوِير [tatˤwiir] development	153.89	11	1102.79	1	2	452.77	0	834	762
تَبَنِّي [tabanniy] adoption	31.52	25	1428.21	4	2	58.91	0	693	368
ساوَمَ [saawama] to compromise	0.08	11	11.78	2	6	24.57	1	729	558
راكَمَ [raakama] to accumulate	0.26	10	58.1	4	6	53.79	3	641	469
ناضَلَ [naadˤala] to struggle	0.83	7	95.34	3	4	28.77	1	662	503
سَيطَرَة [saytˤarah] control	34.75	4	337.11	0	4	336.68	0	668	357
غادَرَ [ɣaadara] to leave	28.84	14	225.31	0	4	186.47	0	646	571
فَلْسَفة [falsafah] philosophy	23.3	7	173.76	0	6	152.84	0	710	448
مُكَثَّف	9.57	12	170.02	0	5	94.99	0	961	378

[mukaθθaf] Condensed									
مُخَيَّم [muxayyam] camp	10.77	11	257.61	1	3	68.52	0	741	641
مُرَشّح [muraʃʃaħ] candidate	37.71	11	518.74	0	6	347	0	907	382
مُشاكَلَة [muʃaakalah] resemblance	0.16	30	3934.07	11	7	7.6	1	845	686
مُشَرِّع [muʃarriʕ] legislator	0.73	32	2501.17	0	4	23.15	0	644	559
مُجامعة [mudʒaamaʕah] sexual intercourse	0.05	41	6902.46	19	3	5.62	0	882	522
مُحَصَّل [muħasˤsˤal] obtained	0.16	16	1069.75	7	5	24.1	1	876	728
زاوَجَ [zaawadʒa] to couple, to pair	0.13	26	544.43	12	5	8.51	3	879	734
تَعارَفَ [taʕaarafa] to know one another	1.3	31	2735.71	13	4	15.61	1	788	585
تَعاظَمَ [taʕaaðˤama] to intensify	4.27	23	623.92	7	3	20.18	0	797	458
تَطابُق [tatˤaabuq] identification	4.94	23	712.59	7	7	31.32	0	811	451
تَسَوُّق [tasawwuq] shopping	2.05	28	1358.12	5	6	19.6	2	729	632
بَيْضَوِيّ [baydˤawiyy] oval	0.13	24	252.37	9	4	40.86	2	771	547
تَخَلّصَ [taxallasˤa] to dispense with	3.98	23	303.87	4	6	77.36	1	864	603
تَقَلّب	3.8	21	420.08	0	6	47.1	0	840	729

[taqallub] turning over									
مُعَوَّض [muʕawwadˤ] compensated	7.33	22	239.91	3	4	27.51	1	1150	977
مُجَلّد [mudʒallad] volume	2.52	27	95.31	0	5	17.22	0	895	782
تَماسَكَ [tamaasaka] to hold together	5.49	27	274.09	2	3	40.86	2	859	586
قِيامة [qiyaamah] resurrection	1.59	43	5132.33	2	2	449.81	0	693	562
كَماليّات [kamaaliyyaat] luxuries, nonessentials	0.31	23	1084.33	7	2	145.8	0	1175	541
قَضى [qadˤaa] to judge	12.09	28	1103.01	5	3	375.48	2	437	296
حَرَّكَ [ħarraka] to move	2.13	12	1294.16	5	6	393.37	3	642	455
ناسَبَ [naasaba] to suit	0.08	24	4216.78	11	7	482.04	3	639	467
تَألّفَ [taʔallafa] to consist of	0.65	32	1556.78	5	5	240.86	2	776	566
شُعوبيَّة [ʃuʕuubiyyah] all nation	0.21	26	1581.62	9	3	201.53	1	1060	572
مُتَفَرِّق [mutafarriq] separated	0.21	43	2169.44	5	7	49.77	1	989	867
تَعَرَّضَ [taʕarradˤa] to be exposed to	91.47	47	2973.11	1	6	400.33	0	659	483
مُخالَفة [muxaalafah] violation	20.65	35	1891.48	3	6	149.74	0	793	602
تَحَقّقَ [taħaqqaqa] to come true	98.94	29	3941.03	5	6	758.26	1	762	556

مُباشَرَة [mubaaʃarah] directly	150.62	32	997.55	1	4	437.25	1	756	459
مُساعَدَة [musaaʕadah] help	48.38	27	2115.4	1	6	746.42	0	795	493
تُرجُمان [turdʒumaan] interpreter	0.31	5	193.58	2	1	4.24	0	945	606
دَسيسة [dasiisah] intrigue	0.08	7	5.24	2	1	0.26	0	736	358
نُعاس [nuʕaas] drowsiness	0.55	8	3.97	0	1	2.93	0	871	706
حُمْق [ħumq] stupidity	0.23	12	24.58	1	1	1.94	0	673	575
شَمْع [ʃamʕ] wax	0.34	10	121.97	0	3	10.03	0	614	540
ثَدْي [θady] bosom	0.47	3	4.43	0	1	4.14	0	519	273
مُشافَهَة [muʃaafahah] orally	0.1	7	9.31	2	3	0.89	0	861	524
راجِمة [raadʒimah] rocket launcher	0.16	6	35.06	0	2	2.99	0	694	522
زِمَام [zimaam] reins of power	9.42	4	34.54	0	1	19.56	0	784	415
رَصِيف [rasˤiif] sidewalk	5.41	11	32.32	0	1	19.86	0	857	694
تَقِيّ [taqiyy] pious	12.43	6	23.6	0	1	17.29	0	548	548
فَتِيل [fatiil] fuse	6.22	9	10.48	0	2	8.94	0	796	719
شَبَح [ʃabaħ] ghost	7.23	2	24.64	0	1	15.12	0	731	575
مُداوَلة	0.26	16	5722.49	7	7	154.49	2	884	717

[mudaawalah] Deliberation									
مُنَوَّع [munawwaʕ] varied	0.13	12	749.01	6	6	118.32	4	716	616
مُمَوَّل [mumawwal] financed	1.33	13	1105.75	2	5	55.59	1	785	696
تأَسَّسَ [taʔassasa] to be established	5.51	15	2297.14	5	6	917.89	3	716	575
مُعَسْكَر [muʕaskar] camp	29.78	8	1142.54	0	8	1142.5	1	697	470
ساهَمَ [saahama] to contribute	22.73	11	906.04	2	4	382.65	1	762	559
مُلاحَظَة [mulaaħaðˤah] notice	17.58	12	549.96	1	4	263.76	1	695	559
تَأَكَّدَ [taʔakkada] to ascertain	12.74	9	1703.43	3	7	1511.9	2	737	581
نَارِيّ [naariyy] fiery	1.98	37	587.82	2	2	216.12	1	690	494
نازِح [naaziħ] emigrant	1.14	8	46.22	0	1	26.43	0	927	418
مِثلِيَّة [miθlyyah homosextuality	0.18	40	5184.02	16	3	984.94	0	848	391
مَبِيع [mabiiʕ] sale, selling	0.65	21	487.59	1	2	122.17	0	524	469
مَغارِب [maɣaarib] west areas	0.03	31	1706.26	6	2	21.24	0	791	559
دَبَّاب [dabbaab] buggy	0.18	12	83.59	0	2	49.08	1	961	820
عادِم [ʕaadim] Exhaust	0.44	14	1028.05	8	1	2.04	0	747	628

بُور [buwr] Uncultivated	3.59	35	145.9	0	2	7.63	0	695	695
حَزْم [ħazm] determination	4.92	14	124.31	1	1	21.17	0	642	522

Appendix C
Items of Study II

Jordanian Colloquial Arabic Words	Modern Standard Arabic Words
برواز	أبكَم
[birwaaz]	[ʔabkam]
Frame	Deaf
بشكير	احتيال
[baʃkiir]	[ʔiħtyaal]
Towel	Fraud
خربطه	أفعى
[xarbatˤah]	[ʔafʕaa]
Chaos	Snake
دغري	ألَم
[duɣrii]	[ʔalam]
Straight	Pain
زَعَل	تَرميم
[zaʕal]	[tarmiim]
Anger	Renovation
سرسري	تَلَف
[sarsarii]	[talaf]
Scoundrel	Damage
طَشت	جدار
[tˤuʃt]	[dʒidaar]
Washtub	Wall
سُفره	زُكام
[sufrah]	[zukaam]
Dinning-table	Cold
سوالف	سُرور
[sawaalif]	[suruur]
Gossip	Happiness
شطه	سَفيه
[ʃatˤtˤah]	[safiih]
Hot Sauce	Silly
كندره	سياج
[kundarah]	[siyaadʒ]
Shoes	Fence
خِتيار	صُحُف
[xityaar]	[sˤuħuf]
Shoes	Newspapers
مناكير	طبيب
[manaakiir]	[tˤabiib]
Nail Polish	Physician
نَصاحَه	عَبَث
[nasˤaaħah]	[ʕabaθ]
	Uselessly

		Obesity
تكسي	[taksi]	Taxi
أسَكِر	[ʔasakkir]	I close
حرامي	ħaraami	Thief
شاكوش	[ʃaakuuʃ]	Hammer
شُرت	[ʃurt]	Power Failure
غلط	[ɣalatˤ]	Mistake
كراج	[karaadʒ]	Garage
كلسات	[kalsaat]	Socks
مُشاجَرَه	[muʃaadʒarah]	Fight
مُنحَدَر	munħadar	Slope
أغادر	[ʔaɣaadir]	I leave
مُنعَطَف	[munʕatˤaf]	Curve
هاتف	[haatif]	Telephone
وِشاح	[wiʃaaħ]	Scarf
إبِل	[ʔibil]	Camel
سَبّاك	[sabaak]	Plumber

The Carrier Sentences

1- هاذ واحد جَبان بعدو بخاف من مُنعَطَف\حرامي

This is a coward as he’s still afraid of a curve/ thief.

2- أثناء تجواله مرَّ الشاب بجانب حرامي\ مُنعَطَف

While the young guy was hanging around, he passed by a thief/ curve.

3- إفتح الدُرج وشيل لِغراض رح تلاقي رَسمِة أفعى\ شاكوش

Open the drawer and remove the stuff, you’ll find a drawing of a snake/hammer.

4- أخرجَ الرجلُ المُسِن من صندوقِهِ شاكوش\ أفعى

The old man has taken a hammer/snake out of his box.

5- أنا إتأكدِت ميه بالميه انو في عندو سَوالِف\ هاتف

I was certain that he had some gossip/a telephone.

6- لهذا السبب تحديداً أصبحت والدتُهُ مُتَضايقه من أيَّة هاتف\ سَوالِف

For this reason, in particular, his mother has become upset about any telephone/gossip.

7- في إبجنب البنك ساحه كبيره بعديها بيجي مُنحَدَر\ كراج

Beside the bank there is a large lot, after which there is a slope/garage.

8- قالَ المُحامي بأنَّ سيارتَهُ قد تعطلت بالقرب من كراج\ مُنحَدَر

The lawyer said that his car had broken down near a garage/slope.

9- لا تفكر ولا يخطر إبالك إني اليوم بدي أَسَكِر\ أغادر

Donʔt think that today I am going to close/leave.

10- حسناً لا تقلق كثيراً لأنني سوفَ أنتظرُكَ قبلَ أن أُغادر\ أَسَكِر

Well do not worry too much as I will wait for you before I leave/close.

11- إبعد هاي الأشياء لهناك مشان ما يصير تلف\ شُرت

Take these things away to avoid any damage/power failure.

12- أخيراً استطاع التقرير أن يُثبتَ أنَّ هذهِ الحالة ناتجه عن شُرت\ تلف

Eventually the report has proven that this is a result of power failure/damage.

13- أنا متأكد إنو هاذ المسؤول ما إنحَط عَبَث\ دُغري

Iʔm sure this person-in-charge wasn’t appointed straight (directly)/uselessly.

14- تيقنتُ بأنَّ سالماً أمضى نِصفَ حياتِهِ دُغري\ عَبَث

I was certain that Salim spent half of his life uselessly/straightforward.

15- أكره ما علي إني أشوف قدامي سِياج\ زَعَل

The worst thing is to see a fence/anger in front of me.

16- سوفَ يُساعدكَ هذا الشيء على التخلص من أيِّ زَعَل\ سِياج

This thing will help you to get rid of any anger/fence.

17- يا عمي شفتلك إنوما في فايده بلبلد لأنها مليانه خربطه\ احتيال

I think the country isn't good as it's full of chaos/fraud.

18- منذُ تأسيسها لم تشهد الدولة حالات احتيال\ خربطه

Since its establishment, the country has not experienced any cases of fraud/chaos.

19- إذا بِدِك خلينا إنزبط نفس اللون على وِشاح\ كُندره

If you want let's match the same color with a scarf/a pair of shoes.

20- لم تمتلك عبير في يومٍ من الأيام ثمن كُندره\ وِشاح

Abeer has never afforded a pair of shoes/a scarf.

21- غريب إنهم يكتبو خبر كامل عن تَكسي\ مُشاجره

It's unusual to write an entire newspaper article about a taxi/fight.

22- نعم لقد قالَ لي بأنَّكَ تستطيع مُشاهَدَتَهُ أينما تُشاهد مُشاجره\ تَكسي

Yes, he told me that you can see it wherever you see a fight/taxi.

23- إذا بدَك نصيحتي لا تفكر فيه وحاول تنسى أي أَلَم\ غَلَط

If you want my advice, don't think about it and try to forget any pain/mistake.

24- حمدت ربي كثيراً لأنه لم يكن هُنالِك غَلَط\ أَلَم

I thanked God much because there was no mistake/pain.

25- بَعدَ طول إنتظار إقتَرَحَ مُصطَفى وزوجَتُهُ إضافة طَبيب\ شَطَّه

After waiting for a long time, Mustafa and his wife suggested adding a physician/hot sauce.

26- بعد ما رِجعَت لقت سحر عندو بالمحل شَطَّه\ طَبيب

When she came back, Sahar found hot sauce/a physician in his shop.

27- ليسَ صعباً أن يتعاملَ إبنُك معَ شخص سرسري\ أبكم

It is not hard for your son to deal with a bad person\deaf.

28- بعديها باسبوعين عرفو إنو الموظف كان أبكم\ سرسري

After two weeks they knew that the employee was deaf/a bad person.

29- غادرَ أسامة غُرفَتَه إلى المطبخ وكان مَعَهُ صُحُف\ بشكير

Osama has left his room to the kitchen carrying newspapers/a towel.

30- كان نايم بَس رمى أحمد عليه من فوق السِدّه بشكير\ صُحُف

He was sleeping when Ahmad threw towel/newspapers at him over the attic.

31- حتى أنهُ لا يعرف ما معنى كَلِمَة سَطِل\ سُرور

Even he does not know the meaning of the word bucket/happiness.

32- بس اعرفي انو مش بسهوله بتقدري تلاقي سُرور\ طُشت

Just I want you to know that you can’t find happiness/a washtub/easily.

33- لا أعتقد بأنَّ هذا الأمر سيؤدي إلى نَصاحه\ زُكام

I do not think that matter will cause obesity/a cold.

34- يا ريت يا ستي وقفت الشغله على شوية زُكام\ نَصاحه

I wish it had been just a matter of cold/obesity.

35- لستُ متأكداً فيما اذا كانت هذهِ الرائحه رائحة إبل\ كلسات

I am not sure whether this is a smell of camels/socks.

36- خلص بِكَفي رح أَنزل هون لأنو هاذ الشارع مليان كلسات\ إبل

That’s enough! I’ll get off here because this street is full of socks/camels.

37- لا يَبدو أنَّ هذا الشيء يُشبِه أيَّ ترميم\ بِرواز

It does not seem that this thing looks like any renovation/frame.

38- اذا بِدِّك تُخلُص مِنها واتجيبلك سعر كويس سويلها بِرواز\ ترميم

If you want to sell it for a good price, make a frame/do renovation for it.

39- لم تَقتَنِع بأنَّ القَضيَّة ليست قضية إزالة مناكير\ جدار

She was not convinced that the issue was not removing some nail polish/a wall.

40- تعال اطَلَّع بالله عليك في هون آثار جدار؟\ مناكير؟

Please come and see if there are any cues for a wall/nail polish here?

41- لقد أخبرتُكَ مراراً وتكراراً أنهُ زَبونٌ سَفيه\ خِتيار

I have told you more than once that he is a silly/an old-man customer.

42- حاول تختصر مشان ما يوخذو عنك فكره إنك خِتيار\ سَفيه

Try to avoid doing this as you don’t want them to think you?re an old man/silly.

43- سوفَ يكون في تلكَ المنطقه المُجاوره سَبَّاك\ سُفره

There will be a plumber/dinning-table in that nearby area.

44- بدي اياك اتفكر قبل ما تجيب سُفره\ سَبَّاك

You should think well before getting a dinning-table/plumber.

45- روح بَس دير بالك كثير بجوز تلاقي مُنعَطَف\حرامي

You can go, but take care as you might find a curve/thief.

46- أُقسِم بأنَّهُ قد تكاد تخلو هذهِ المدينة من أيِّ حرامي\ مُنعَطَف

I swear that there is almost no thief/curve in this city.

47- قعدو عمال السرك أكثر من ساعتين بدورو على أفعى\ شاكوش

The circus workers spent more than two hours looking for a snake/hammer.

48- أنظر هناكَ لهذا الشيء يبدو كأنهُ شاكوش\ أفعى

Look there! That thing looks like a hammer/snake.

49- لا الشرطه ولا القاضي بقدرو يسجنو على سوالِف\ هاتف

Neither the police nor the judge can send people to prison for gossip/a telephone.

50- يعتقد الأخوان بأن أصل الخلاف مُجرَّد هاتف\ سوالِف

The brothers think that the debate has developed from a mere telephone call/gossip.

51- عيب ما بصير تقول للناس إتفضلو وبعدين إتقعدهم على مُنحَدَر\ كراج

Shame on you! It is inappropriate to tell people “come in” and then seat them on a slope/garage roof.

52- حاولي أن تَجِدي موقعاً آخر بعيداً عن أيِّ كراج\ مُنحَدَر

Try to find another location away from any garage/slope.

53- والله لو بتحط إيدَك إبإيد مين ما رَح أَسَكِر\ أغادِر

I swear by God that I’ll never close/leave.

54- بعدَ أن عَرَفتُ قرار الحكومه لم يكن بوسعي إلا أن أغادِر\ أَسَكِر

After I had learnt about the government decision, I could not do anything but to leave/close.

55- ما رح إصير مشكلة اذا صار فيه تلف\ شُرت

There will be no problem if a damage/power failure has happened.

56- لا أظنكم تُبالون أبداًحتى لو أدى ذلك إلى شُرت\ تلف

I do not think that you care even if this causes a power failure/damage.

57- الأسبوع الماضي ما كانو بانينها دُغري\ عَبَث

Last week they didn’t build it straight/uselessly.

58- كانت الفتاه لا ترغَب ولا تُحِب أن تُجَرِّبَها عَبَث\ دُغري

The girl did not want to try it uselessly/straight.

59- لحديت الآّن مش قادر أصدق إنو كل هاذ سِياج\ زَعَل

I still can’t believe that all of this is a fence/anger.

60- فَكِّر مره أخرى لأنه لن يَنفَعَكَ أيٌّ زَعَل\ سِياج

Think again because anger/a fence is not good for you.

61- صدقني يا سيدي انوالجيران إعتبروا إلي صار خربطه\ احتيال

Believe me, sir, the neighbors considered what has happened chaos/fraud.

62- تستَطيع أن تتوقع ذلك وتتجنب ما يمكن أن يَحدُث من احتيال\ خربطه

You can expect this and avoid what might happen of fraud/chaos.

63- لا تخافي لأنو ما رَح أرَوِّح قبل ما ألاقيلِك وِشاح\ كُندره

Don’t worry! I’m not going to go home before finding a scarf/a pair of shoes for you.

64- شاهَدَ محمدٌ عَمَّهُ وهو يحرق كُندره\ وِشاح

Mohammad saw his uncle burning a pair of shoes/a scarf.

65- بطَّل يرجع مرة ثانيه بسبب تَكسي\ مُشاجره

He decided not to come back again because of a taxi/fight.

66- لقد فَقدَ الشاب محفظتَهُ ليلةَ أمس في مُشاجره\ تَكسي

Last night the young man lost his wallet in a fight/taxi.

67- حاولت اكثير بس كل يوم كنت أحس إنو هاذ أَلَم\ غلط

I tried a lot, but day after day I felt that was a pain/mistake.

68- لقد أرشدتني عِدَّت مرات كيف يمكن أن أتجنب أيَّ غلط\ أَلَم

You have taught me how to avoid any mistake/pain.

69- هذا المُنتَج الذي يُباع في الأسواق هوَ عِباره عن مُنتَج طَبيب\ شَطَّه

This product sold in markets is a physician’s/hot sauc product

70- إذا بدك ترتاح وتريحُه روح جيبلو شَطَّه\ طَبيب

If you want to please yourself and him, bring him hot sauce/a physician.

71- كانَت المجموعه تهتم لأمرِهذا الشخصٍ لأنَّهُ سرسري\ أبكَم

The group cared about this person because he was a bad one/deaf.

72- دير بالك إزل لسانك واتجيب سيره إنو أبكَم\ سرسري

Be careful! Don’t forget and disclose the secret that he’s deaf/a bad person.

73- لقد حضرَ ونسيَّ حاجاتَه لذلك طَلَبَ مني أن أُحِضرَ لَهُ صُحُف\ بشكير

He has come and forgot his stuff, so he asked me to bring him some newspapers/a towel.

74- الشَغلِه سهله كثير هات كيس وحُط فيه بشكير\ صُحُف

It is very easy: bring a bag and put a towel/some newspapers in it.

75- لقد تَمَنَّيتَ أن يكونَ عندَ أصدقائِكَ سَطِل\ سُرور

You wished that your friends had a washtub/happiness.

76- عُمرَك ما رح إدَّخِل على بيتَك سُرور\ طُشت

You will never bring happiness/a bucket to your home.

77- كل هذه المشاكل التي تعانيها هي مشاكل نصاحه\ زُكام

All of these problems that you have are natural consequences of obesity/cold.

78- ما حدا بالدنيا كلها بحب اصير عندو زُكام\ نصاحه

None in this world likes to have a cold/obesity.

79- لقد ادعى مِراراً وتكراراً أنهُ تاجر صَغيّر يبيع إبِل\ كَلسات

He often claimed that he was a small dealer selling camels/socks.

80- إتصدقو انبَسَطِت إكثير لأنو أول مره بَجَرِّب هيك كَلسات\ إبِل

Believe or not! I was very happy because it was the first time I try such these socks (on)/camels.

81- يعتقد أحمد بِأنَّ شَكلَها سيكون جميل من دون تَرميم\ بِرواز

Ahmad thinks that it would look nice without renovation/ a frame.

82- ما رَح تِقدَر إتبَلِّش مِن دون بِرواز\ تَرميم

You can’t start without a frame/renovation.

83- طَلَبَ خالدٌ من خطيّبَتِهِ أن تَضَعَ مناكير\ جدار

Khaled asked his fiancée to put (apply/build) some nail polish/a wall.

84- عُمرَك إبحياتَك شُفِت مَي مِجَّمّعَه على جدار\ مناكير

Have you ever seen water gathering on a wall/ nail polish.

85- اتفقَ جميع سكان الحي على أنَّ حارس البيت شخصٌ سَفيه\ خِتيار

The people of the quarter have agreed that the janitor is a silly person/an old man.

86- ما بحبو لأنو بِظَل إعايرني إني خِتيار\ سَفيه

I don’t like him because he keeps telling me I’m an old man/silly.

87- كَانَ الجَد يُخَطِط ويُدَبِر لإحضار سَبَّاك\ سُفره

The grandfather was planning and arranging to get a plumber/dinning-table.

88- يا سيدي ما عليك لأنك ابتقدر إتلاقي أي سُفره\ سَبَّاك

Don’t worry man! You can find any dining-room/plumber.

Appendix D
Items of Study III

MSA prime/target	JCA prime/target	Frequency	Unrelated MSA Prime	Unrelated JCA Prime
حَقيبَة	شَنْتِة		نافِذَة	شُبَّاك
ħaqiibah	ʃanteh	11.89	naafiðah	ʃubaak
bag	bag		window	window
أَفْعَى	حَيِّة		قِطّة	بِسِّة
ʔafʕaa	ħayyeh	0.49	qitˤtˤah	bisseh
snake	snake		pussycat	pussycat
شُرْفَة	بَلّكونِة		قُبلَة	بوسِة
ʃurfah	balkoneh	4.08	qublah	boseh
balcony	balcony		kiss	kiss
مَسْحُوق	بودرة		لَكْمَة	بُكس
masħuuq	bodrah	1.77	lakmah	buks
powder	powder		fist	fist
سِتَارَة	بُردايِّة		وِشَاح	شَال
sitaarah	burdaayyeh	1.09	wiʃaaħ	ʃaal
curtain	curtain		scarf	scarf
مِعطَف	جاكيت		أنبوب	مَاسورَة
miʕtˤaf	dʒaaket	0.75	ʔunbuub	maasuurah
coat	coat		pipe	pipe
رَجُل	زَلَمِة		هاتِف	تَلَفُون
radʒul	zalameh	112.18	haatif	talafon
man	man		telephone	telephone
فُنْدُق	أوتيل		مِقْوَد	ستيرنج
funduq	ʔutel	51.78	miqwad	stiiring
hotel	hotel		steering wheel	steering wheel
نُقُود	مَصَاري		ثَدي	بِز
nuquud	masˤaarii	2.42	θadii	biz
money	money		bosom	bosom
إِطَار	بِروَاز		خُرْطَوم	بَربيش
ʔitˤaar	birwaaz	0.94	xurtˤuum	barbiiʃ
frame	frame		hose	hose
طَمَاطِم	بَنْدورَة		حِذَاء	كُنْدَرَة
tˤamaatˤim	bandorah	0.26	ħiðaaʔ	kundarah
tomato	tomato		shoes	shoes
مُسْتَقِيم	دُغْرِي		مُسِن	خِتيَار
mustaqiym	duɣrii	1.92	musin	xityaar
straight	straight		elderly	elderly
مِطّرَقَة	شَاكُوش	1.33	مِدفَأة	صوبَة
mitˤraqa	ʃaakuuʃ		midfaʔah	Sobah

hammer	hammer		Heater	Heater
حَسَاء	شورَبَة		سُعَال	قَحَّة
ħasaaʔ	ʃuurabah	0.31	suʕaal	gaħħah
soup	soup		cough	cough
مَصْعَد	أصَنْصيل		رَصِيف	بَنكيت
masˤʕad	ʔasˤansˤel	0.57	rasˤiif	banket
elevator	elevator		sidewalk	sidewalk
قُمَامَة	زبالِة		بَعُوض	قارِص
qumaamah	zbaaleh	1.77	baʕuudˤ	gaarisˤ
garbage	garbage		mosquito	mosquito
حَاسُوب	كُمبيوتَر		وِسَادَة	مَخَدِّة
ħaasuub	kumbyuutar	0.23	wisaadah	maxaddeh
computer	computer		pillow	pillow
بُنْدُقِيَّة	بَارودِة		مِذيَاع	رَاديو
bunduqiyyah	baaruudeh	4.01	miðyaaʕ	raadio
rifle	rifle		radio	radio
قُبَّعَة	طَاقِيِّة		قِلادَة	سِنْسَال
qubbaʕa	tˤaagyyeh	1.82	qilaadah	sinsaal
hat	hat		necklace	necklace
مِنْدِيل	مَحْرَمَة		مائِدَة	سُفْرَة
mindiil	maħramah	0.52	maaʔidah	sufrah
tissue	tissue		dining table	dining table
نُزْهَة	طَشِّة		بَرَاعَة	شَطَارَة
nuzhah	tˤaʃʃeh	3.07	baraaʕah	ʃatˤaarah
excursion	excursion		skilfulness	skilfulness
عَمَل	شُغُل		مُشَاجَرَة	هوشِة
ʕamal	ʃuɣul	304.17	muʃaaɟarah	hoʃeh
work	work		fight	fight
فَقْر	طَفَر		اكتِظَاظ	عَجقَة
faqr	tˤafar	4.21	ʔiktiðˤaaðˤ	ʕadʒgah
poverty	poverty		overcrowding	overcrowding
حِمْيَة	روجِيم		كَذِب	عَرْط
ħimyah	rodʒiim	0	kaðib	ʕartˤ
diet	diet		lying	lying
هَنيئاً	صِحتين		هَيَّا	يَلّى
haniiʔan	sˤiħten	0.65	hayyaa	yalaa
good appetite	good appetite		let’s go	let’s go
حَر	شوب		ألَم	وَجَع
ħar	ʃob	14.33	ʔalam	wadʒaʕ
hot	hot		pain	pain
أحسَنت	برافو		إصْمِت	انتشَب
ʔaħsant	bravo	2.42	ʔisˤmit	ʔintʃab
bravo	bravo		shut up	shut up
إنجَاب	خِلفِة		تَرميم	تَشطيب
ʔindʒaab	xilfeh	1.63	tarmiim	taʃtˤiib
giving birth	giving birth		renovation	renovation

غَضَب	زَعَل		إهمَال	تَطنيش
ɣadˤab	zaʕal	16.91	ʔihmaal	tatˤniiʃ
anger	anger		neglect	neglect
إغلَاق	تَسكير		غَداً	بُكْرَة
ʔiɣlaaq	taskiir	0.21	ɣadan	bukrah
closure	closure		tomorrow	tomorrow
خَطَأ	غَلَط		خَارِج	بَرَّة
xatˤaʔ	ɣalatˤ	50.04	dʒxaari	barrah
wrong	wrong		outside	outside
تَفَقّد	تَشييك		بَحْث	دوارَة
tafaqqud	taʃiiyyk	18.49	baħθ	dwarah
check	check		search	search
لِأجل	مِشَان		أيضَاً	كَمَان
liʔadʒl	miʃaan	7.88	ʔaydˤan	kamaan
for the sake of	for the sake of		also	also
تَبَاهِي	فَشْخَرَة		ضَجَر	زَهَق
tabaahiy	faʃxarah	0.39	dˤadʒar	zahag
showing off	showing off		boredom	boredom
شَراهَة	فَجْعَنِة		رَكلَة	شوتِة
ʃaraahah	fadʒʕaneh	0.26	raklah	ʃoteh
gluttony	gluttony		kick	kick
مَجَّاناً	بَلاش		جَفَاف	نَشَافِة
madʒdʒaanan	balaaʃ	8.5	dʒafaaf	naʃaafeh
free	free		dryness	dryness
مَكابِح	بريكات		آلَة	مَاكينَة
makaabiħ	brekaat	0.05	[ʔaalah]	maakiinah
brakes	brakes		machine	machine
مُشَاهَدَة	فُرْجِة		تَوبِيخ	بَهدَلِة
muʃaahadah	furdʒeh	11.57	tawbiix	bahdaleh
watching	watching		telling off	telling off
لِص	حَرَامي		سَائِق	شوفير
lisˤ	ħaraamii	1.38	saaʔiq	ʃufer
thief	thief		driver	driver
قَليلاً	شوَيِّة		جَيّد	كوَيِّس
qaliilan	ʃwayyeh	44.68	dʒayyid	kwayyis
little	little		good	good

Appendix E
Items of Study IV with Their Characteristics

MSA /ħ/	Frequency	Neighbors	UP	Duration	JCA /ħ/	Frequency	Neighbors	UP	Duration
سُحُب [suħub] clouds	5.55	2	439	531	مَحْرَمَة [maħramah] tissue	6	5	436	520
تَحَدُّث [taħadduθ] talking	5.55	4	631	783	كَحَش [kaħaʃ] to dismisse	5.36	11	457	621
مَسحُوق [masħuuq] powder	4.64	12	574	667	سُحليّة [suħlyyeh] lizard	4.55	4	314	756
اقتِحام [ʔiqtiħaam] breaking into	4.64	2	353	770	حَفَرْتَلِي [ħafartalii] low-class person	3.82	0	651	848
حَفَاوَة [ħafaawah] welcoming	3.73	4	608	800	حَنَفِيِّة [ħanafyyeh] faucet	6.32	0	531	859
حُقْنَة [ħuqnah] shot	5	4	371	544	حَاووز [ħaawwz] reservoir	3.64	0	509	875
حَقِيبَة [ħaqiibah] bag	5.41	6	538	731	شَبَّاح [ʃabbaaħ] trousers strap	5.05	8	894	1028
مُبَاح [mubaaħ] permissible	5.41	6	775	775	تَشْليّح [taʃliiħ] robbery	5.45	15	811	967
اجتياح [ʔidʒtiyaaħ] invasion	4.86	4	800	938	مكَرْسَح [mkarsaħ] handicapped	5.59	2	463	932
نُبَاح [nubaaħ] barking	4.5	2	359	774	حوش [ħoʃ] backyard	5	9	595	730
حَفْنَة [ħafnah] handful	3.36	6	409	640	مْسَحوِل [msaħwil] dragging on the ground	4.77	2	450	715
حَانَ [ħaana] bar	5.27	6	562	677	أفْكَح [ʔafkaħ] waddle	4.14	8	559	707

مَكابِح [makaabiħ] brakes	4.36	3	745	912	شَطَح [ʃatˁaħ] to exaggerate	3.86	13	529	707
مِصْبَاح [misˁbaaħ] lamp	5.23	0	279	867	نَاصِح [naasˁiħ] overweight	6.55	11	659	817
نِكَاح [nikaaħ] matrimony	4.86	1	648	783	شَحَّاطة [ʃaħħaatˁah] flip-flops	4.5	2	593	815
انحِدار [ʔinħidaar] descending	5.05	2	317	747	حَرّوش [ħarruuʃ] yellow melon	3.14	1	442	870
وِشَاح [wiʃaaħ] scarf	3.82	0	647	784	طَاحَش [tˁaaħaʃ] to crowd	2.95	8	575	726
رِيَاح [riyaaħ] winds	5.95	6	630	778	كَحْتُوت [kaħtuut] stingy	5.09	0	438	810
شُح [ʃuħ] scarsity	4.09	13	446	630	سَنْكُوح [sankuuħ] vulger	4.05	0	582	985
أطّرُوحَة [ʔutˁruuħah] dissertation	3.64	0	525	777	بَحَش [baħaʃ] to dig	4.95	12	571	749
حُشُود [ħuʃuud] crowds	4.09	1	754	901	مْقَحْمِش [mgaħmiʃ] crispy	5.5	2	414	864
رَحِب [raħib] spacious	4.27	8	430	539	مْدَحْبَر [mdaħbar] rounded	5.32	2	405	688
صَحِيفَة [sˁaħiifah] newspaper	5.27	3	500	716	دَواحِل [dawaaħil] marbles	5.23	1	488	879
استِحسَان [ʔistiħsaan] admiration	4.55	0	705	1014	حِز [ħiz] split	5.73	26	675	675
استَحضَرَ [ʔistaħdˁara] to recall	4.05	1	554	884	حَرْدانِة [ħardaaneh] angry-wife	5.68	6	342	829
تَحَفّظ [taħaffuðˁ] reservation	4.91	1	641	750	جَحَر [dʒaħar] to stare at	6.05	14	435	534
مُتَاح [mutaaħ] available	5.82	4	685	826	حَشْران [ħaʃraan] congested	6.27	3	526	876

مُستَحَثّات [mustaħaθθaat] fossils	2.5	0	634	1365	رَدّاحَة [raddaaħah] scolding woman	4.64	5	683	901
جَامِح [dʒaamiħ] crazy	2.77	6	630	757	مْحَالاة [mħaalaah] showing off	4.64	2	575	769
تَنْقِيح [tanqiiħ] review	3.68	7	715	862	حُزّيرَة [ħuzzerah] puzzle	5.82	2	479	796
مُسْتَحْضَر [mustaħdˤar] formulation	4.77	2	566	849	شُحبَار [ʃuħbaar] soot	5.73	0	348	864
ضَرِيح [dˤariiħ] tomb	4.23	6	568	708	شَحطَة [ʃaħtˤah] a little bit	5.55	11	562	717
صَرْح [sˤarħ] edifice	4.27	11	507	679	حومرَة [ħomrah] lipstick	5.82	2	427	607
إِفْسَاح [ʔifsaaħ] making room for	3.82	3	776	923	مَدْحُوش [madħuuʃ] tightly packed	5.32	4	673	835
مُحْتَال [muħtaal] swindler	5.36	5	758	881	حَرامي [ħaraamii] thief	6.18	4	614	750
قاحِلَة [qaaħilah] arid.feminine	4.5	7	375	608	شَرَاشيح [ʃaraaʃiiħ] vulgers	4.18	1	551	918
شَرِيحَة [ʃariiħah] slice, section	5.91	11	515	764	شَرحَة [ʃarħah] small box	4.23	11	465	600
فاحِشَة [faaħiʃah] fornication	4.73	1	575	823	مبارِح [mbaariħ] yesterday	6.45	4	690	824
تَمحِيص [tamħiisˤ] testing	3.86	5	854	1060	مقَحِّف [mgaħħif] shrewd	4.59	7	636	764
حِنْطَة [ħintˤah] wheat	3.64	3	383	615	كَحَّة [kaħħah] cough	6.77	6	367	529

MSA /l/	Frequency	Neighbors	UP	Duration	JCA /l/	Frequency	Neighbors	UP	Duration
لَدْغَة [ladɣah] sting	3.91	0	279	495	كَلسُون [kalson] Underwear	5.59	0	408	652
لَيسَ [laysa] a negative particle	5.32	2	324	546	بَلْكونِة [balkoneh] Balcony	6	2	425	735
لَذيذ [laðiið] delicious	5.5	0	578	728	تَشْلِيح [taʃliiħ] Robbery	5.41	15	658	784
لَبيب [labiib] intelligent	4.05	5	319	744	مَسطُول [mastˤuul] thick-headed	4.77	7	615	686
لائِحَة [laaʔiħah] sign	5.36	2	400	613	لَفْحَة [lafħah] Scarf	5.64	8	262	508
لَكَمَ [lakama] to punch	4	7	357	532	لَطْخَة [latˤxah] Stupid	4.5	3	273	527
مَسْلُوب [masluub] robbed	3.45	18	803	929	شَال [ʃaal] Headscarf	5.5	43	553	553
قِلادَة [qilaadah] necklace	4.23	3	287	647	بِيل [biil] Flashlight	3.64	19	500	500
حَليف [ħaliif] ally	4.91	11	680	841	ريالِة [ryaaleh] Slaver	5.41	0	520	648
تَحْلِيق [taħliiq] soaring	4.82	8	573	687	لِسَّة [lissah] not yet	6.5	2	277	508
تِجْوَال [tidʒwaal] walking around	3.09	1	528	940	فانيلَة [faanelah] Undershirt	5.68	1	353	703
جَعَلَ [dʒaʕala] to make	5.05	8	392	603	بُكْلِة [bukleh] hair clip	6.14	2	265	444
أرسَلَ [ʔarsala] to send	5.36	1	389	558	بْلوزِة [bluuzeh] Blouse	5.68	1	347	559

وَسائِل wasaaʔil methods	6.09	2	594	698	سَجلونِة [saɈloneh] Chesterfield	3.5	1	316	766
أطْفَال [ʔatˤfaal] children	5.95	4	416	740	لَخْمِة [laxmeh] empty-headed	4.41	9	377	590
مَنْزِل [manzil] house	5.73	6	746	746	بَالطُو [baaltˤo] Coat	4.18	0	417	628
مَحْصُول [maħsˤuul] crop	4.95	7	737	887	مْشَلِّط [mʃalliT] rude/pimp	4	8	524	601
لَعِين [laʕiin] cursed	3.95	1	511	635	دَلِّة [dalleh] coffee flusk	4.95	19	417	551
طَلِيق [tˤaliiq] free	4.55	7	543	677	بَلّاعَة [ballaaʕah] Sewer	4.45	2	549	748
تَبَلْوَرَ [tabalwara] to crystallize	3.95	0	510	708	شِلِن [ʃilin] five piasters	5.82	2	386	456
مُلصَق [mulsˤaq] sticker	4.41	1	337	651	مْليح [mliiħ] Good	6.23	4	688	688
مُطْلَق [mutˤlaq] absolute	4.91	2	393	646	شَلّوت [ʃaluut] Kick	5.55	0	681	810
بَلْدَة [baldah] town	5.36	5	378	602	شْوال [ʃwaal] Bag	5.86	5	583	651
بُلْعُوم [bulʕuum] pharynx	3.5	0	409	707	شِلِّة [ʃilleh] Gang	5.45	6	494	607
فَيلَق [faylaq] corps	3	2	399	772	لَطَش [latˤaʃ] to steal	5	13	429	565
لَحْد [laħd] sepulcher	2.59	4	415	567	بَلَف [balaf] to steal	3.91	15	420	544
لَدُود [laduud] mortal enemy	4.18	1	307	625	مَلْط [maltˤ] Nakedness	3.27	7	391	491
سَلِيط [saliitˤ] sharp-tongued	3.55	4	592	693	زَلْط [zaltˤ] Nakedness	3	7	442	551
تَلْبِيَة	4.64	3	476	711	جِل	5.91	19	532	532

[talbyah] compliance					[dʒil] Gel				
حُقُول [ħuquul] fields	4.64	5	644	750	فِلْدِة [fildeh] military Jacket	4.05	2	324	491
جَمْيل [dʒamiil] beautiful	5.95	11	591	726	شَلاطيف [ʃalaatˁiif] thick lips	4.27	0	289	849
تَقْلِيم [taqliim] clipping, triming	4.5	15	657	787	بَلَهَموطي [balhamuutˁii] Greedy	3.86	0	352	790
جَلِيّ [dʒaliyy] obvious	4.09	6	655	655	بَلاش [balaaʃ] Free	6.27	4	530	686
مُعْتَل [muʕtal] sick	3.86	7	645	645	مْنَيَّل [mnayyal] Poor	4.68	3	624	714
تَمَلّق [tamalluq] hypocrisy	3.09	8	627	763	جْلاطَة [dʒlaatˁah] Mud	3.77	3	259	619
أمْوَال [ʔamwaal] money.plural	6.32	9	748	892	لاطَة [laatˁah] Dumb	3.91	8	422	579
أعَالَ [ʔaʕaala] to provide for	3.42	11	604	690	بَرطِيل [bartˁiil] hush money	2.53	3	483	763
أخْلَى [ʔaxlaa] evacuate	5.26	17	424	559	صَنْدَل [sˁandal] Sandals	5.79	0	396	608
مُتَسَوِّل [mutasawwil] beggar	4.26	5	548	895	دْرِل [dril] Drill	4.95	0	155	462
تَوَسُّل [tawassul] to entreat	4.79	4	639	727	هيلَمِة Helameh	4.47	1	386	574

MSA /i/	Frequency	Neighbors	UP	Duration	JCA /i/	Frequency	Neighbors	UP	Duration
صَائِب [sˤaaʔib] right	4.37	14	572	657	بِز [biz] Bosom	5.05	21	516	516
مِقْوَد [miqwad] steering wheel	4.32	3	342	626	لِز [liz] come closer	4.21	27	514	514
وِعَاء [wiʕaaʔ] container	4.37	0	302	628	زِنْار [zinnaar] Belt	5.58	0	430	707
حِذَاء [ħiðaaʔ] shoes	4.84	1	357	681	جِزْدَان [dʒizdaan] Wallet	5.84	1	291	725
مِزْمَار [mizmaar] flute	4.26	3	251	630	بِس [bis] Tomcat	6.32	18	725	645
طِرَاز [tˤiraaz] model	4.58	0	657	657	زِفت [zift] showing disapproval	6.32	4	451	562
طَماطِم [tˤamaatˤim] tomato	4.42	0	366	615	مِشَان [miʃaan] for the sake of	5.53	2	547	640
بَاهِظ [baahiðˤ] costly	4.26	3	484	543	بَلِّش [balliʃ] Start	5.58	4	568	724
مُعْضِلَة [muʕdˤilah] problem	4.53	1	284	605	مَزْمِز [mazmiz] enjoy eating or drinking	4.32	0	395	635
غِبْطَة [ɣibtˤah] jealousy	2.68	0	336	536	مْصَربِع [msˤarbiʕ] Impatient	4	2	434	652
فَارِه [faarih] luxury	4	28	583	672	دَبِل [dabil] Double	5.32	5	388	462
مُلائِم [mulaaʔim] suitable	4.68	5	489	692	دوزِن [dozin] tune, adjust	4.37	2	301	531
حائِل	3.21	17	591	690	بِرواز	4.95	1	286	831

[ħaaʔil] obstacle					[birwaaz] Frame				
تَتِمَّة [tatimmah] continuation	3.53	0	329	613	مسَختِن [msaxtin] Tipsy	3.74	1	477	742
مِقبَض [miqbadˤ] handle	4	4	342	608	سْرِنج [srindʒ] Shot	5.16	1	371	691
بارِقَة [baariqah] a glimpse of hope	2.79	14	395	613	مِس [mis] female teacher	5.79	14	538	538
آنِف [ʔaanif] preceding	3.58	3	528	660	طَقِع [tˤagiʕ] Awesome	5	9	275	379
مَائِدَة [maaʔidah] dining table	4.42	12	343	576	جِم [dʒim] Gym	5.63	20	482	482
بائِد [baaʔid] extinct	3.32	15	537	673	سِشوَار [siʃwaar] hair dryer	5.79	2	245	798
مُرتَعِب [murtaʕib] frightened	4.32	4	639	756	سِكْرَاب [sikraab] junk yard	5.42	0	382	821
طَلَائِع [tˤalaaʔiʕ] forefronts	3.37	1	437	712	فرِش [friʃ] Fresh	5.63	6	451	639
مُبْهِر [mubhir] amazing	4	7	537	606	بَراطِم [baraatˤim] Lips	4.26	1	392	668
فَرائِس [faraaʔis] preys	3.89	3	768	921	داشِر [daaʃir] Bastard	5.05	14	352	606
قَادِم [qaadim] (up)coming	4.95	11	511	612	خَشِم [xaʃim] Nose	5.95	7	428	495
غِرَار [ɣiraar] pattern, example	3.89	7	551	618					
حِرْفَة [ħirfah] handicraft	4.42	4	448	688					
مُبْرِح [mubriħ] Sever	3.47	4	555	717					

مِضْمَار [midˤmaar] racetrack	3.79	3	414	707
قَراصِنَة [qaraasˤinah] pirates	4.84	0	523	714
فارِس [faaris] horseman	4.74	20	661	807
مِحْرَاب [miħraab] niche	4.16	2	483	1000
سِتَار [sitaar] screen	4.58	3	481	829
سِلْعَة [silʕah] a piece of merchandise	4.37	4	300	619
إِغْدَاق [ʔiɣdaaq] giving generously	2.79	3	229	768
جَوَارِب [dʒawaarib] socks	5	5	717	821
لاذِع [laaðiʕ] blunt	4.74	1	321	572
طَرائِف [tˤaraaʔif] anecdotes	4.32	1	338	724
دَامِغ [daamiɣ] irrefutable	3.16	1	526	603
إِبَاحَة [ʔibaaħah] permission	5.32	4	397	630
دَواجِن [dawaadʒin] poultry	4.79	1	487	862

MSA /u/	Frequency	Neighbors	UP	Duration	JCA /u/	Frequency	Neighbors	UP	Duration
بُرَاز [buraaz] excrement	4.16	2	618	754	رُزْنَامَة [ruznaamah] Calendar	5.47	0	287	684
دُمْيَة [dumyah] doll	4.37	1	319	548	بُرْدَايِّة [burdaayeh] Curtain	6.05	0	478	689
مُسِن [musin] elderly	4.95	4	494	644	زْقُرت [zgurt] reliable person	4.79	1	226	740
شُرْفَة [ʃurfah] balcony	4.74	5	385	620	كُنْدَرَة [kundarah] Shoes	6.11	0	335	589
فُكَاهَة [fukaahah] joke	4.58	0	425	800	خُشِّة [xuʃʃeh] very small room	5.05	3	543	698
كُرَة [kurah] ball	5.32	10	262	389	بُكْلِة [bukleh] hair clip	5.68	3	326	543
نُضْج [nudˤdʒ] maturity	4.74	0	532	636	هُلّيلة [hulleleh] big celebration	3.79	0	378	669
تُخمَة [tuxmah] indigestion	4.21	1	352	484	صُرمْاية [sˤurmaayeh] old shoes	5.05	0	426	758
كُمَّثْرَى [kummaθraa] pear	3.95	0	312	715	بُكس [buks] box (punch)	5.37	6	544	544
زُلال [zulaal] white of egg	3.95	1	363	640	بُكسِة [bukseh] box(container)	4.84	4	415	535
قُبَّعَة [qubbaʕah] hat	4.16	0	357	547	كُشِّة [kuʃʃeh] Forelock	5	2	413	546
مُحْتَضَر [muħtadˤar] dying	3.89	6	455	724	كُبَاية [kubaayeh] Cup	4.21	2	438	647
تَمَعُّن [tamaʕʕun] scrutiny	4.68	2	319	661	طُشْت [tˤuʃt] Washtub	5.32	3	377	483

فُلْك [fulk] ship	4.63	3	397	522	مُزِّة [muzzeh] attractive woman	6.05	5	378	564
دُعَابَة [duʕaabah] joking	4.47	1	429	592	بُسْطَار [busˤtˤaar] military boots	4.74	0	256	676
عُصْبَة [ʕusˤbah] league, group	4.42	7	316	504	قُرْنِة [gurneh] Cushion	5.21	2	345	581
مُعَبَّد [muʕabbad] paved	4.53	7	706	901	طُرْ [tˤur] Dismiss	3.68	30	333	333
مُضْطَرِب [mudˤtˤarib] confused	4.26	0	580	679	مُغّيطَة [muɣɣetˤah] Rubber	4.42	0	318	663
مُفَدَّى [mufaddaa] dearest	4.16	1	360	591	حُزّيرَة [ħuzzerah] Puzzle	5.37	0	443	748
تَقادُم [taqaadum] prescription	3.58	2	424	654	طُقّيعَة [tˤuggeʕah] Inferior	4.53	0	353	735
تَهَكُّم [tahakkum] sarcasm	3.74	2	341	631	قُرّيطَة [gurretˤah] hair’s breadth	4	1	638	784
تَقَاعُس [taqaaʕus] slackenin	4.26	2	649	771	بُرنيطَة [burnetˤah] Hat	3.53	0	350	741
تَبَدُّد [tabaddud] dispersion	3.74	9	578	686	مُقْليعَة [mugleʕah] Sling	3.53	1	366	741
سُعَال [suʕaal] cough	5.21	1	636	730	زُر [zur] to accelerate	4.16	32	508	508
رُعَاع [ruʕaaʕ] vulger	2.47	4	514	554	ثُم [θum] Mouth	5.58	19	455	455
قُبَالَة [qubaalah] opposite to	3.63	3	420	606	جُغْمِة [dʒuɣmeh] Bite	4	0	322	571
تَبَرُّج [tabarrudʒ] primping	4.79	13	658	767	بُدِي [budii] Body	4.11	2	409	576
قُمَامَة [qumaamah]	5.05	1	401	595	طُرطيرَة [tˤurtˤerah]	4.37	0	340	711

rubbish					Rickshaw				
قُصَارَى [quSaaraa] utmost	4.68	2	449	663	كُرْكَعَة [kurkaʕah] Turtle	3.84	0	332	603
حُطَام [ħutˤaam] wrechage	4.58	0	388	695	زُمْيرَة [zumerah] Pipe	4.95	0	379	733
رُفَات [rufaat] mortal remains	3.84	3	403	873	لَزيق [luzzeg] adhesive tape	5.89	2	464	810
تَشاجُر [taʃaadʒur] fight	5.05	2	487	786	دُش [duʃ] Shower	6.32	14	460	645
تَعَاقُب [taʕaaqub] succession	4.68	5	478	713	نُص [nusˤ] Half	5.79	17	650	650
تَغَلغُل [taɣalɣul] penetration	3.84	1	442	771	طَمْايِّة [tˤumaayeh] hide-and-seek	4.74	1	337	751
حُقْبَة [ħuqbah] era	3.84	5	366	621	فُرْجِة [furdʒeh] Show	5.26	3	448	581
كُثْبَان [kuθbaan] dunes	4.32	0	323	902	دُغْري [duɣrii] Straight	6.16	2	298	521
زُكَام [zukaam] flu	4.74	1	410	724	شُرت [ʃurt] Shorts	5.68	6	415	534
تَصَاعُد [tasˤaaʕud] ascending	4.63	6	469	854	شُفير [ʃufer] Driver	5.47	0	412	668
صَلُحَ [sˤaluħa] to be pious	4.79	2	341	575	شُبَّاك [ʃubbaak] Window	5.95	0	745	830
عَظُمَ [ʕaðˤuma] to be great	3.89	0	407	498	شُبّيبِة [ʃubbebeh] Pipe	4.16	1	450	784

MSA Filler / ħ/	JCA Filler /ħ/	MSA Filler /l/	JCA Filler /l/
تَكْرِيم [takriim] honoring	بوزَة [buuZah] ice cream	تَكْرِيم [takriim] Honoring	بوزَة [buuZah] ice cream
مُثابَرَة [muθaabarah] perseverance	كَرت [kart] card	مُثابَرَة [muθaabarah] Perseverance	كَرت [kart] card
هاتِف [haatif] telephone	بَاص [baasˤ] bus	هاتِف [haatif] Telephone	بَاص [baasˤ] bus
بَدانَة [badaanah] obesity	كروز [kroz] a carton of cigarettes	بَدانَة [badaanah] Obesity	كروز [kroz] a carton of cigarettes
أفعَى [ʔafʕaa] snake	فَشَر [faʃar] shut up!	أفعَى [ʔafʕaa] Snake	فَشَر [faʃar] shut up!
تَنَزَّهَ [tanazzaha] to go on a picnic	طَبّون tˤabbon [trunk]	تَنَزَّهَ [tanazzaha] to go on a picnic	طَبّون [tˤabbon] trunk
مِرآب [mirʔaab] garage	فَطّرَز fatˤtˤraz [to backflip]	مِرآب [mirʔaab] Garage	فَطّرَز [fatˤtˤraz] to backflip
شِفاء [ʃifaaʔ] cure, recovery	عَنفَص [ʕanfasˤ] to explode with anger	شِفاء [ʃifaaʔ] cure, recovery	عَنفَص [ʕanfasˤ] to explode with anger
صَعَدَ [sˤaʕada] climb	بَكْرَج [bakradʒ] coffee pot	صَعَدَ [sˤaʕada] Climb	بَكْرَج [bakradʒ] coffee pot
نَافِذَة [naafiðah] window	بَاكور [baakuur] cane	نَافِذَة [naafiðah] Window	بَاكور [baakuur] cane
أعَادَ [ʔaʕaada] to return	بَرَّاد [barraad] refrigerator	أعَادَ [ʔaʕaada] to return	بَرَّاد [barraad] refrigerator
غَدَاً [ɣadan] tomorrow	طَفَر [tˤafar] poverty	غَدَاً [ɣadan] Tomorrow	طَفَر [tˤafar] poverty
رَفيق [rafiiq] companion	فَاع [faaʕ] to spread out	رَفيق [rafiiq] Companion	فَاع [faaʕ] to spread out
سُقُوط [suquutˤ] falling	كَتْكُوت [katkuut] chick	سُقُوط [suquutˤ] Falling	كَتْكُوت [katkuut] chick

طَفْيف	[tˤafiif]	slight
إطَار	[ʔitˤaar]	frame
فَريد	[fariid]	unique
نَكَثَ	[nakaθa]	to violate
اقتَرَفَ	[ʔiqtarafa]	to commit
مُنذُ	[munðu]	since
فارِس	[faaris]	horseman
سِتَار	[sitaar]	screen
بَاكِراً	[baakiran]	early
سِلْعَة	[silʕah]	a piece of merchandise
إغْدَاق	[ʔiɣdaaq]	giving generously
صَائِب	[sˤaaʔib]	right
مِقْوَد	[miqwad]	steering wheel
وِعَاء	[wiʕaaʔ]	container
مِزْمَار	[mizmaar]	flute

كْمَاج	[kmaadʒ]	pita bread
بَنْج	[bandʒ]	anesthetic
جَك	[dʒak]	jack
مَفزور	[mafzuur]	leaking
جَكَر	[dʒakar]	stubbornness
شَاكووش	[ʃaakuuʃ]	hammer
بَلْكونِة	[balkoneh]	balcony
مرَطْرِط	[mratˤritˤ]	slack
مَسطُول	mastˤuul	[thick-headed]
لَطْخَة	[latˤxah]	stupid
شَال	[ʃaal]	scarf
بيِل	[biil]	flashlight
ريالِة	[ryaaleh]	slaver
مَريول	[maryuul]	coveralls
عْيال	[ʕyaal]	kids

طَفْيف	[tˤafiif]	Slight
إطَار	[ʔitˤaar]	Frame
فَريد	[fariid]	Unique
نَكَثَ	[nakaθa]	to violate
اقتَرَفَ	[ʔiqtarafa]	to commit
مُنذُ	[munðu]	Since
بُرَاز	[buraaz]	Excrement
دُمْيَة	[dumyah]	Doll
مُسِن	[musin]	Elderly
شُرْفَة	[ʃurfah]	Balcony
فُكَاهَة	fukaahah	joke
كُرَة	kurah	ball
نُضْج	nudˤdʒ	maturity
تُخمَة	tuxmah	indigestion
كُمَّثْرَى	kummaθraa	pear

كْمَاج	[kmaadʒ]	pita bread
بَنْج	[bandʒ]	anesthetic
جَك	[dʒak]	jack
مَفزور	[mafzuur]	leaking
جَكَر	[dʒakar]	stubbornness
شَاكووش	[ʃaakuuʃ]	hammer
تحَرْكَش	[tħarkaʃ]	to provoke
مَحْرَمَة	[maħramah]	tissue
كَحَش	[kaħaʃ]	dismiss
حَنَفِيِّة	[ħanafiyyeh]	faucet
حَاووز	[ħaawuuz]	reservoir
شَبَّاح	[ʃabbaaħ]	trousers strap
مكَرْسَح	[mkarsaħ]	handicapped
حوش	[ħoʃ]	backyard
أفْكَح	[ʔafkaħ]	waddle

طِرَاز	[tˤiraaz]	model
طَماطِم	[tˤamaatˤim]	tomato
بَاهِظ	[baahiðˤ]	costly
مُعْضِلَة	[muʕdˤilah]	problem
غِبْطَة	[ɣibtˤah]	jealousy
فَارِه	[faarih]	luxury
لَيسَ	[laysa]	a negative particle
لَذيذ	[laðiið]	delicious
لَبيب	[labiib]	intelligent
تِجْوَال	[tidʒwaal]	walking around
جَعَلَ	[dʒaʕala]	to make
لِسَّة	[lissah]	not yet
فانيلَة	[faanelah]	undershirt
عَوَاطلي	[ʕawaatˤlii]	jobless
جَلافِة	[dʒalaafeh]	roughness
دَرْبِيل	[darbiil]	binoculars
بَالِة	[baaleh]	used-clothes shop
سِشوَار	[siʃwaar]	hair dryer
تَنَكِة	[tanakeh]	tin can
سِكْرَاب	[sikraab]	junk yard
كَرَتِة	[karateh]	shoe-horn
فَرْدِة	[fardeh]	one of a pair
قُبَّعَة	qubbaʕah	hat
دُعَابَة	duʕaabah	joke
عُصْبَة	ʕusˤbah	league,group
مُعَبَّد	muʕabbad	paved
مُضْطَرِب	[muDtˤarib]	Confused
تَقادُم	[taqaadum]	Prescription
نِكَاح	[nikaaħ]	Matrimony
انحِدار	[ʔinħidaar]	Descending
وِشَاح	[wiʃaaħ]	Scarf
رِيَاح	[riyaaħ]	Winds
استِحسَان	[ʔistiħsaan]	Admiration
شَطَح	[ʃatˤaħ]	to exaggerate
نَاصِح	[naasˤiħ]	overweight
رُزْنَامَة	[ruznaamah]	calendar
بُرْدَايِّة	[burdaayeh]	curtain
زْقُرت	[zgurt]	reliable person
كُنْدَرَة	[kundarah]	shoes
خُشِّة	[xuʃʃeh]	very small room
فُرْجِة	[furJeh]	show
دُغْري	[duɣrii]	straight
صُرمْاية	[sˤurmaayeh]	old shoes
بُكس	[buks]	box (punch)

MSA Filler /i/	JCA Filler /i/	MSA Filler /u/	JCA Filler /u/
جُدْرَان [dʒudraan] walls	كَبُوت [kabuut] coat	جِدَار [dʒidaar] Wall	دَفَش [dafaʃ] to push
تَقَصَّى [taqasˤsˤaa] to investigate	دَفَش [dafaʃ] to push	تَقَصَّى [taqasˤsˤaa] to investigate	طَعَج [tˤaʕadʒ] to bend
مَحْروقَات [maħruuqaat] fuel	طَعَج [tˤaʕadʒ] to stab	زَائِف [zaaʔif] False	طَنْط [tˤantˤ] sissy
وَقُود [waquud] fuel	طَنْط [tˤantˤ] sissy	آزَرَ [ʔaazara] to support	زَعْلان [zaʕllan] upset
رُبَّان [rubbaan] captain	زَعْلان [zaʕllan] upset	خَاضَ [xaadˤa] to engage in	دَرابزين [daraabziin] handrail
جَسَد [dʒasad] body	حُومرَة [ħomrah] lipstick	إغلَاق [ʔiɣlaaq] Closing	كَشْخَة [kaʃxah] showing off
أَثْنَاء [ʔaθnaaʔ] during	كَشْخَة [kaʃxah] showing off	جَسَد [dʒasad] Body	فَشَخ [faʃax] to hurt someone’s head
أَمْس [ʔams] yesterday	فَشَخ [faʃax] to hurt someoneʔs head	أَثْنَاء [ʔaθnaaʔ] During	كَسْتَك [kastak] strap
أَصْدَرَ [ʔasˤdara] to release	كَسْتَك [kastak] strap	أَمْس [ʔams] Yesterday	مْكَرْبِج [mkarbidʒ] dry and rigid
زَاوَلَ [zaawala] to practice	كَرْبُوج [karbuudʒ] cute	أَصْدَرَ [ʔasˤdara] to release	ستْيرِنج [stiiring] steering wheel
الآن [ʔalʔaan] now	سَكْسُوكَة [saksuukah] goatee	الآن [ʔalʔaan] Now	شْراك [ʃraak] a type of thin bread
شَنَّ [ʃanna] to launch an attack	شْراك [ʃraak] a type of thin bread	شَنَّ [ʃanna] to launch an attack	سَرْسَرَة [sarsarah] swindling
غَادَرَ [ɣaadara] to leave	سَرْسَرَة [sarsarah] swindling	غَادَرَ [ɣaadara] to leave	بَلّكَش [balkaʃ] to tamper with

آخَرْ [ʔaaxar] another	بَلْكَش [balkaʃ] to tamper with	آخَرْ [ʔaaxar] Another	قَزَّز [gazzazz] to annoy
أسْفَرَ [ʔasfara] to result in	جَرسون [garson] waiter	أسْفَرَ [ʔasfara] to result in	جَخ [dʒax] to spoil himself with
رُبَّمَا [rubamaa] perhaps	قَزَّز [gazzazz] to annoy	مَناقِب [manaaqib] Virtues	طَنْجَرَة [tˤandʒarah] cooking pot
هَزَمَ [hazama] to defeat	جَخ [dʒax] to spoil himself with	هَزَمَ [hazama] to defeat	عَشَان [ʕaʃaan] because of
أرَّقَ [ʔarraqa] to keep someone awake	طَنْجَرَة [tˤandʒarah] cooking pot	أرَّقَ [ʔarraqa] to keep someone awake	هيلَمِة [helameh] deception
أمَام [ʔamaam] in fron of	عَشَان [ʕaʃaan] for the sake of	أمَام [ʔamaam] in front of	هيطَلِيِّة [hetˤaliyyeh] a type of sweets
ذَهَبَ [ðahaba] to go	بُسْطَار [bustˤaar] military shoes	ذَهَبَ [ðahaba] to go	مْنَيَّل [mnayyal] poor
صَلُحَ [sˤaluħa] to be pious	ثُم [θum] mouth	آنِف [ʔaanif] Preceding	مْليح [mliiħ] good
طُرْفَة [tˤurfah] joke	دُش [duʃ] shower	بائِد [baaʔid] Extinct	مَلْط [maltˤ] naked
عَظُمَ [ʕaḌuma] to become great	زُر [zur] accelerate	بَائِع [baaʔiʕ] salesman, seller	مْلَخْلَح [mlaħlaħ] street-smart
فُلْك [fulk] ship	شُبَّاك [ʃubbaak] window	بارِقَة [baariqah] a glimpse of hope	مَقْلَمِة [maglameh] pen case
قُبَالَة [qubaalat] opposite to		بَاطِن [baatˤin] Core	مْشَلَّط [mʃallitˤ] rude,pimp
قُصَارَى [qusˤaaraa] utmost		تَتِمَّة [tatimmah] Continuation	لاطَة [laatˤah] dumb
قُمَامَة [qumaamah] rubbish		رَائِج [raaʔidʒ] common	كْلَشِن [klaʃin] Kalashnikov

			rifle
كُثْبَان [kuθbaan] dunes		غِرَار [ɣiraar] Pattern	صَنْدَل [sˤandal] sandals
مُحْتَضَر [muħtadˤar] dying		مِضْمَار [midˤmaar] Racetrack	شِلِن [ʃilin] five piasters
مُسْتَحْضَر [mustaħdˤar] formulation		قَادِم [qaadim] (up)coming	مْسَختِن [msaxtin] very relaxed
مُستَحَثَّات [mustaħaθθaat] fossils		لاذِع [laaðiʕ] Blunt	مِس [mis] female teacher
مُرَجَّح [muradʒdʒaħ] likely		مَائِدَة [maaʔidah] dinning table	مَزْمِز [mazmiz] enjoy eating or drinking
مُحْتَال [muħtaal] swindler		مِحْرَاب [miħraab] Niche	مْرَطرِط [mratˤritˤ] slack
مُتَاح [mutaaħ] available		بَلْدَة [baldah] Town	مْرَبْرِب [mrabrib] chubby
مُبَاح [mubaaħ] permissible		جَلاء [dʒalaaʔ] Clarity	مَخَدِّة [maxaddeh] pillow
قَدَح [qadaħ] goblet		جَمْيل [dʒamiil] Beautiful	كَنْزِة [kanzeh] sweater
فَحْوَى [faħwaa] essence, main content		لائِحَة [laaʔiħah] List	قِب [gib] go away
صَفَحَ [sˤafaħa] to forgive		لَدْغَة [ladɣah] Bite	فْرِش [friʃ] fresh
سُحْت [suħt] ill-gotten property		لَذيذ [laðiið] Delicious	فَارْدِه [faardeh] wedding procession
سُحُب [suħub] clouds		مَنْزِل [manzil] House	مِشَان [miʃaan] for the sake of

Appendix F
Instructions and the Consent Form

The English Translation for the Frequency-rating Instructions

A Side-experiment

Part I: You will listen to a number of Jordanian Arabic words. After each word, you will be asked to rate how frequent the word is used by Jordanians on a 7-point scale. Press play to listen to the word.

1= not frequent 2 3 4= moderately frequent 5 6 7= very frequent.

Part II: You will listen to a number of Standard Arabic words. After each word, you will rate, on a 7-point scale, how frequent the Standard word is used in the contexts of Modern Standard Arabic (in newspapers, books, news, documentaries, interviews, etc).

1= not frequent 2 3 4= moderately frequent 5 6 7= very frequent.

اختبار تكرار الكلمات

الجزء الأول: سوف تستمع إلى عدد من الكلمات المستعملة باللهجه العربيه الأردنيه. بعد الاستماع لكل كلمة سوف تقوم بتقييم تلك الكلمه من حيث درجة تكرارها (شيوعها) ككلمة مستعمله. انقر على زر التشغيل لتستمع إلى الكلمة.

كثيرة الشيوع =7 6 5 متوسطة الشيوع = 4 3 2 قليلة الشيوع =1

الجزء الثاني: سوف تستمع إلى عدد من الكلمات المستعملة باللغه العربيه الفصحى. بعد الاستماع لكل كلمة سوف تقوم بتقييم تلك الكلمه من حيث درجة تكرارها (شيوعها) في الأوساط التي تُستَخدَم بها اللغه الفصحى (في الصحف والكتب ونشرات الأخبار والبرامج الوثائقية والمقابلات التلفزيونية...الخ) الانقر على زر التشغيل لتستمع إلى الكلمة.

كثيرة الشيوع=7 6 5 متوسطة الشيوع = 4 3 2 قليلة الشيوع =1

The English Translation for the Word-association Instructions
A side- experiment

You will listen to 40 Modern Standard Arabic words. Then you will be asked to suggest, for each Standard Arabic word, its best translation equivalent in Jordanian Arabic.

Press play to listen to the word.
The best Jordanian Arabic equivalent for this word is ------------------------

تجربة ارتباط الكلمات

سوف تستمع إلى ٤٠ كلمة من كلمات اللغة العربية الفصحى. بعد الاستماع إلى كل كلمة سوف تقوم باقتراح أفضل مرادف (ترجمة) لهذه الكلمة باللهجه العربيه الأردنيه
أفضل كلمة أردنية دارجة مرادفة (مُشابهة) لهذه الكلمة هي ....

The English Translation for the Informed Consent (the main experiments)

Informed Consent Form (The Main Experiments)

Study Name: Auditory Word Recognition in Arabic (Diglossia)

Principal Investigator: Moh'd Al-Omari

Research Supervisor: Prof. Kevin Russell

Affiliation: University of Manitoba, MB, Canada.

This consent form, a copy of which you may keep for your records and reference at this time, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, you should feel free to contact us. Please take the time to read this carefully and to understand any accompanying information.

Purpose of the study

Mr. Moh'd Al-Omari is conducting this study for his doctoral dissertation in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Linguistics, under the supervision of Prof. Kevin Russell. The purpose of this research is to investigate how fast and accurate native speakers of Jordanian Arabic recognize Arabic words in the spoken modality. We estimate it will take no longer than **-25 minutes** to complete the experiment.

Subjects Selection

You have been chosen to take part in this project because you match the criteria that the researcher is looking for. That is, you are a native speaker of Jordanian Arabic with at least 12 years of formal education in Arabic.

Procedures

***** You will be hearing a number of individual words. Sometimes the word is a real Arabic word, a meaningful Arabic word. Sometimes the word is not a real Arabic word, a non-sense (meaningless) word. **Press the YES button if the stimulus is a real word of Arabic. Press the NO button if it is a non-sense (meaningless) word. You need to give your decision as fast and as accurately as possible. (**To appear on the consent form of the morphology experiment)

***** You will listen to Arabic sentences ended with words or non-words. **Press the YES button if the sentence ended with a word in any forms of Arabic (Jordanian Arabic or Standard Arabic). Press the NO button if the sentence ended with a non-word. You need to give your decision as fast and as accurately as possible. There will be a two-second interval between your response and the next sentence. (**To appear on the consent form of the code-switching experiment)

***** You will listen to a list of Arabic words in pairs. In each pair, the first word is always a real Arabic word. The second word, however, can be a real Arabic word or a non-sense (meaningless) word**. Press the YES button if the second word of each pair is a real word of any form of Arabic (Jordanian Arabic or Standard Arabic). Press the NO button if it is a non-sense (meaningless) word. You need to give your decision as fast and as accurately as possible. (**To appear on the consent form of the priming experiment)

* This experiment consists of two sections. In both sections, you will listen to a number of individual words in both Standard and Jordanian varieties of Arabic. Your task is to

decide whether the introduced word has a specific predetermined sound in it. **Press the YES button if you hear the sound in the word, otherwise you don?t have to press anything until the next word appears. (**To appear on the consent form of sound-monitoring experiment)

Withdrawal

You can withdraw from the experiment anytime by pressing the escape key on the computer keyboard. If you decide to withdraw from the experiment before its completion, you will not receive any compensation.

Confidentiality

The research data contains no personal identifiers and, thus, poses no risk of identification to participants. The computer will only identify participants by a numeric code — and there will be no record anywhere that ties the numeric codes to participants' identities. The computer will be recording the participants' answers to each stimulus and the time (in milliseconds) that it took them to respond. This anonymous data will never be destroyed and may be made public (e.g., disseminated to academic occasions such as journals, conferences and/or posted electronically online).

Potential risks and discomforts

There is no more than minimal risk to individuals who participate in this research study.

Benefit

A potential benefit of participating, other than receiving the compensation, is that you may learn something about the psycholinguistic factors that could determine how fast you recognize Arabic words. You might also get an answer to the question of how Arabic words are stored and processed in the lexicon (i.e. mental dictionary)

Feedback

The researcher will provide the participants with two types of feedback. The first feedback will be a summary of the research purpose, hypothesis and methods. The participants will receive this feedback written on a sheet of paper immediately after the experiment. The second feedback will be about the results and findings of the study. The researcher will post the results on-line at *thelinguisticresearch.com* after analysing the data statistically. The results feedback will be available to the participants by 06/2016.

Dissemination

The experiment results will be disseminated to the participants, the researcher's supervisor (Dr Russell) and to the research committee members. The results could also be disseminated to academic occasions such as journals, conferences, etc.

Compensation

Each participant will be awarded 5 JD (about 10 CAD) for his/her participation.

Researcher: Moh'd Ahmad Al-Omari Tel: [redacted]
E-mail: alomamak@myumanitoba.ca

Supervisor: Prof. Kevin Russell E-mail: Kevin.Russell@ad.umanitoba.ca

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a subject. In no way does this waive your legal rights nor release the researchers, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time, and /or refrain from answering any questions you prefer to omit, without prejudice or consequence. Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation.

The University of Manitoba may look at your records to see that the research is being done in a safe and proper way.

This research has been approved by the Joint-Faculty REB at University of Manitoba. If you have any concerns or complaints about this project you may contact any of the above-named persons or the Human Ethics Coordinator at humanethics@umanitoba.ca.

A copy of this consent form has been given to you to keep for your records and reference.

Participant?s Signature ________________________ Date ____________

قسم اللغويات- جامعة مانيتوبا

نموذج اشعار بالموافقة (التجربة الرئيسية)

عنوان البحث : الإدراك السمعي للكلمات العربية في سياقاتها العامية والفُصحى
اسم الباحث : محمد أحمد العمري
اسم المشرف: د. كِيفن رَسِل

يمثل نموذج الموافقة هذا -والذي سيُعطى المشترك نسخة منه للاحتفاظ بها- جزءاً من عملية الاشعار بالموافقة الكلية . سيقوم هذا النموذج بإعطائك فكرة عامة عن موضوع البحث والأشياء التي ستتضمنها مشاركتك . إذا رغبت بالحصول على توضيح أكثر حول أي مسألة آتية أرجو عدم التردد بالسؤال . الرجاء أخذ الوقت الكامل في قراءة واستيعاب المعلومات التي يتضمنها هذا الطلب .

هدف الدراسة :

سيقوم السيد محمد العمري بإجراء دراسة تحت إشراف الدكتور. كِيفن رَسِل لتكون جزءً من أُطروحته للحصول على درجة الدكتوراه في الدراسات اللغوية.

الهدف من الدراسة هو التعرف على مدى سرعة ودقة إدراك الكلمة العربية المسموعة في سياقاتها العامية والفُصحى.

من المتوقع ان لا تستمر التجربة أكثر من 25 دقيقة تقريباً .

اختيار العينة :

لقد تم إختيارك لكي تشارك في هذه التجربة لأنه قد توفرت لديك معايير التجربة التي يريدها الباحث ، ذلك أن اللهجة العربية الأردنية هي لغتك الأم بالإضافة أنك أنهيت تعليمك الأساسي والثانوي باستخدام اللغة العربية .

إجراءات المشاركة :

- سوف تستمع إلى عدد من الكلمات. في بعض الأحيان ستكون هذِه الكلمات عبارة عن كلمات عًرَبية لها معنى وفي أحيانٍ أخرى ستكون عبارة عن كلمة لا يوجد لها معنى في اللغة. عند الاستماع إلى كُل كلمة الرجاء **أن تُقَرر بأسرع وقت ممكن وبدقة** فيما إذا كانت تلك الكلمةعبارة عن كلمة حقيقية (ذات معنى) أم غير حقيقية (لا معنى لها). الرجاء الضغط على زر **نعم** إذا كانت الكلمة ذات معنى باللغة العربية أو على زر **لا** إذ كانت اللفظة بدون معنى. (التجربة الأولى)

- سوف تقوم بالاستماع إلى جمل باللغة العربية بعضها مُنتهية بكلمات ذات معنى وبعضها مُنتهية بكلمات ليس لها معنى . إضغط على زر **نعم** إذا كانت الكلمة الأخيرة من الجملة لها معنى (سواء أكانت الكلمة عامية أم فصحى). أضغط على زر **لا** إذا كانت الكلمة الأخيرة من الجملة ليس لها معنى . يجب أن تكون إجابتك سريعة ودقيقة بقدر المستطاع . سوف يكون هنالك فاصل ثانيتين بين إجابتك والجملة التالية. (التجربةالثانية)

- سوف تستمع إلى زوج مِن الكلمات (كلمتين متتابعتين) . ستكون الكلمة الأَولى من كل زوج من الكلمات عِبارة عن كلمة عربية لها معنى. ستكون الكلمة الثانية من كل زوج إما كلمة عًرَبية لها معنى أوعبارة عن كلمة لا معنى لها في اللغة العربية. عند الاستماع إلى كل زوج من الكلمات الرجاء **أن تُقَرر بأسرع وقت ممكن وبدقة** فيما إذا كانت الكلمة **الثانية** عبارة عن كلمة حقيقية (ذات معنى) أم غير حقيقية (لا معنى لها). الرجاء الضغط على زر **نعم** إذا كانت الكلمة ذات معنى باللغة العربية ،سواءٍ أكانت الكلمة فصحى أم عامية، أو على زر **لا** إذا كانت الكلمة بدون معنى. (التجربة الثالثة)

- تتألف هذهِ التجرُبَة من شقين. في كلا الشقين ستستمع إلى عدد من الكلمات العربية العامية أو الفُصحى. المَطلوب منك هوأن **تُقَرر بأسرع وقت ممكن وبدقة** في ما إذا كانت هذه الكلمات تحتوي على صوت معين أم لا. الرجاء الضغط على زر **نعم** إذاسمعت الصوت ،الذي سيتم تحديده، في الكلمة المنطوقة. اذا لم تكن الكلمة تحتوي عى هذا الصوت الرجاء الانتظار حيث سيقوم الحاسب تلقائياً بِنُطق الكلمة التالية بعد مرور ثلاث ثوانٍ.(التجربة الرابعة)

الانسحاب من التجربة

يمكنك الانسحاب من التجربة في أي وقت تريد وذلك من خلال الضغط على مفتاح Esc الموجود على لوحة المفاتيح. إذا اخترت الانسحاب من التجربه قبل انتهائها سوف لن تحصل على أية تعويض مالي.

السرية :

لَن يُطلَب مِنك خلال مشاركتك بهذِهِ التجرُبة تزويد الباحث باسمك أو بأي بيانات شخصية عنك يتمكن من خلالها الآخرون من تحديد هويتك. سوف يقوم جهاز الحاسوب باعطاء كل مشترك رقم. وسوف لن يكون هناك أية طريقة تربُط هذا الرقم بهوية المُشارِك. سوف يقوم الحاسوب بتسجيل إجابات المشاركين حَولَ اللفظة المَسموعة وسرعة الإستجابة لتلك اللِفظة بأجزاء الثانية. لن يقوم الباحث باتلاف هذه البيانات مجهولة الصاحب كَما أنها سَوفَ تُعرَض للعلن في المجلات والمؤتمرات العلميةا كما من الممكن أن توضع على الشبكة العنكبوتية.

الأخطار المفترضة :

سوف لن يكون هنالك أية خطر على الأشخاص المشاركين بالتجربة .

الفائدة:

بمشاركتك في هذه التجربة وحصولك على التغذية الراجعة من المتوقع أن تتعلم بعض الأشياء عن العوامل اللغوية/ النفسية المؤثره في سرعة ادراكك للكلمات العربية المسموعة. كما يمكنك معرفة فيما إذا كانت اللغة العربية هي لغة ثانية لمتحدثي العربية أم أنها جزء من لغتنا العربية العامية الأُولى.

التغذية الراجعة :

سيقوم الباحث بتزويد المشاركين في البحث بنوعين من التغذية الراجعة. النوع الأول سيكون عبارة عن ملخص يحتوي على هدف الدراسة وفرضيتها وكيفية اجراءها. سيكون هذا الملخص مكتوباً على ورقة وسيُقَدَّم مباشرة بعد اجراء التجربة. أما النوع الثاني من التغذية فهو عبارة عن ملخص لنتائج الدراسة. سيقوم الباحث بتزويد المشاركين بنتائج البحث بعدَ أن يقوم يتحليل البيانات احصائياً. يستطيع المشارك معرفة نتائج البحث بحلول شهر حزيران من عام 2016 ستكون التغذية الراجعة مُتاحة للمشاركين عَبرَ زيارة الموقع التالي: *thelinguisticresearch.com*

التعويضات والمكافأت :

سَيُمنَح كل مشترك مبلغ خمسَة دَنانيرأُردنية كتعويض له عن المشاركة .

كشف البيانات :

سوف تكشف نتائج هذه التجربة لمشرف البحث د. كِيفن رَسِل وأعضاء هيئة البحث. سوف تُعرَض أيضاً نتائج هذه التجربة في الأوساط الأكاديمية مثل المجالات والمؤتمرات العلمية.

الباحث : محمد أحمد العمري _ ت : [redacted] _ **البريد الالكتروني :** alomamak@myumanitoba.ca

المشرف على البحث : د. كِيفن رَسِل _ **البريد الالكتروني :kevin.russell@ad.umanitoba.ca**

إن توقيعك على هذا النموذج يعني أنك فهمت المعلومات المتضمنة بخصوص مشاركاتك في مشروع البحث ووافقت على المشاركة كأحد أفراد عينة البحث .إن موافقتك على المشاركة لاتعني بأنك تنازلت عن حقوقك القانونية أو أنك أعفيت الباحث أو جهة البحث من حقوقهم القانونية أو المهنية . لك حرية الاختيار بالانسحاب من المشاركة بأي وقت كما أن لك حرية الاختيار بأن لا تعطي أي إجابة عن أية سؤال مطروح دون أي إجحاف أو عواقب . لك الحرية المطلقة بطلب أي توضيح أو معلومات جديدة خلال مشاركتك .

من الممكن أن تقوم جامعة مانيتوبا بالاطلاع على سجل المشترك للتأكُد من أنَّ التجربة قد أُجريت بشكل آمِن ومناسب.

تمت الموافقة على هذا البحث من قبل مجلس أخلاقيات البحوث البشرية في جامعة مانيوتوبا. إذا كان لديك أية استفسار أو شكوى حول هذا المشروع يمكنك الاتصال بإحد الأشخاص المذكورين آأنفاً أومُنسِّقَة أَخلاقيات البحث العلمي على البريد الإلكتروني ca.umanitoba@humanethics

أعطيت لك نسخة من هذا النموذج للإحتفاظ بها .

توقيع المشترك /
التاريخ /
البريد الإلكتروني للمشترك إذا رغب بمعرفة نتائج البحث (اختياري)/

UNIVERSITY
OF MANITOBA
Research Ethics and Compliance
Office of the Vice-President (Research and International)

Human Ethics
208-194 Dafoe Road
Winnipeg, MB
Canada R3T 2N2
Phone +204-474-7122
Fax +204-269-7173

APPROVAL CERTIFICATE

December 8, 2015

TO: **Moh'D Ahmad Al-Omari** **(Advisor K. Russell)**
Principal Investigator

[redacted]

FROM: **Lorna Guse, Chair**
Joint-Faculty Research Ethics Board (JFREB)

Re: **Protocol #J2015:136**
"Auditory Word Recognition in Arabic (diglossia)"

Please be advised that your above-referenced protocol has received human ethics approval by the **Joint-Faculty Research Ethics Board**, which is organized and operates according to the Tri-Council Policy Statement (2). **This approval is valid for one year only**.

Any significant changes of the protocol and/or informed consent form should be reported to the Human Ethics Secretariat in advance of implementation of such changes.

Please note:

- **If you have funds pending human ethics approval, please mail/e-mail/fax (261-0325) a copy of this Approval (identifying the related UM Project Number) to the Research Grants Officer in ORS in order to initiate fund setup. (How to find your UM Project Number: http://umanitoba.ca/research/ors/mrt-faq.html#pr0)**

- **if you have received multi-year funding for this research, responsibility lies with you to apply for and obtain Renewal Approval at the expiry of the initial one-year approval; otherwise the account will be locked.**

The Research Quality Management Office may request to review research documentation from this project to demonstrate compliance with this approved protocol and the University of Manitoba *Ethics of Research Involving Humans.*

The Research Ethics Board requests a final report for your study (available at: http://umanitoba.ca/research/orec/ethics/human_ethics_REB_forms_guidelines.html) **in order to be in compliance with Tri-Council Guidelines.**

umanitoba.ca/research

References

Abu-Rabia, S. (2000). Effects of exposure to literary Arabic on reading comprehension in a diglossic situation. *Reading and Writing, 13*(1-2), 147-157.

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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Albirini, A. (2011). The sociolinguistic functions of codeswitching between standard Arabic and dialectal Arabic. *Language in Society, 40*(5), 537-562.

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40*(1), 41-61.

Al-Khatib, M., & Sabbah, A. E. H. (2008). Language choice in mobile text messages among Jordanian university students. SKY Journal of Linguistics, 21, 37–65.

Allan, K. K. (1982). The development of young children's metalinguistic understanding of the word. *Journal of Educational Research, 76*(2), 89-93.

Almahmoud, M. (2013). Investigating status planning through studying language attitudes. *IPEDR*, 68(11), 53-60.

Al-Sughayer, K. I. (1990). *Aspects of comparative Jordanian and modern standard Arabic phonology* Available from Linguistics and Language Behavior Abstracts (LLBA).

Amara, M. (2006, October). The vitality of the Arabic language in Israel from a sociolinguistic perspective. *Dalah's Newsletter*

Amara, M. (2011). *Arabic language in Israel: Contexts and challenges*. Jordan: Dar Al Fiker. (In Arabic).

Amor P. D. M., & Maad, R. Ben. (2013). The role of Arabic orthographic literacy in the phonological awareness of Tunisian children. *The African Symposium: An Online Journal of the African Educational Research Network, 13*(1), 19-27.

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects. *Journal of Memory and Language, 25*(6), 726-740.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*(5), 255-259.

Arnqvist, A. (1992). The impact of consonant clusters on preschool children's phonemic awareness: A comparison between readers and nonreaders. *Scandinavian Journal of Psychology, 33*(1), 29-35.

Baalbaki, R.H. (2008). *Al-Mawrid: A modern Arabic-English dictionary* (22 ed.). Beirut, Lebanon: Dar El Ilm Lilmalayin.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language, 37*(1), 94-117.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55*(2), 290-313.

Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R.H. Baayen, & R. Schreuder (Eds), *Morphological Structure in Language Processing* (pp. 355-390). Berlin: Mouton de Guyter.

Baayen, R. H., Tweedie, F. J., & Schreuder, R. (2002). The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon. *Brain and Language, 81*(1-3), 55-65.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon, 2*(3), 419-463.

Baetens-Beardsmore, H. (1986). *Bilingualism: Basic principles*. Clevedon, Avon, England: Multilingual Matters.

Badawi, E.S. (1973). *Mustawayat al-arabiyya l-mu'assira fi misr*. Cairo: dar al-ma'arif.

Balling, L. W. (2008): A brief introduction to regression designs and mixed-effects modelling by a recent convert. *Copenhagen Studies in Language*. 36:175-192.

Balling, L. W., & Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes, 23*(7-8), 1159-1190.

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition, 125(1),* 80–106.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283-316.

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy, 9*(1), 95-114.

Bassiouney, R. (2006). *Functions of code switching in Egypt: Evidence from monologues* Brill.

Bat-El, O. (2003). Semitic verb structure within a universal perspective. In J. Shimron (Ed.), *Language processing and language acquisition in a root-based morphology* (pp.29-59). Amsterdam: John Benjamins.

Bates, D. Maechler, M., Bolker, B., & Walker, S. (2016). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8, http://CRAN.R-project.org/package=lme4.

Bauer, L. (2004). *A glossary of morphology*. Washington, DC: Georgetown University Press.

Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *Journal of Memory and Language, 35*(6), 801-820.

Benmamoun, E. (1999). Arabic morphology: The central role of the imperfective. *Lingua, 108*(2-3), 175-201.

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.

Berent, I., Vaknin, V., & Marcus, G. F. (2007). Roots, stems and the universality of lexical representations: Evidence from Hebrew. *Cognition, 104*, 254–286.

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 489-511.

Birdsong, D. (1999). Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 1-21). Lawrence Erlbaum: Mahwah, NJ.

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual Language Control in Perception versus Action: MEG Reveals Comprehension Control Mechanisms in Anterior

Cingulate Cortex and Domain-General Control of Production in Dorsolateral Prefrontal Cortex. J*ournal of Neuroscience, 36*(2), 290-301.

Bolte, J., & Connine, C. M. (2004). Grammatical gender in spoken word recognition in German. *Perception & Psychophysics, 66(6),* 1018-1032.

Boudelaa, Sami. (2014). Is the Arabic mental lexicon morpheme-based or stem-based? Implications for spoken and written word recognition. In E. Saiegh-Haddad, & R. Joshi (Eds.), *Handbook of Arabic literacy: insights and perspectives* (Vol. 9, Literacy Studies) (pp. 31-54). Dordrecht: Springer Netherlands.

Boudelaa, S., & Marslen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing: Evidence from modern standard Arabic. In J. McQueen, & A. Cutler (Eds.), *Proceedings of SWAP (workshop on spoken word access processes (vol.1)* (pp. 23-26). Nijmegen, Netherlands: Max-Planck Institute for Psycholonguistics.

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition, 81*(1), 65-92.

Boudelaa, S., & Marslen-Wilson, W. D. (2004a). Abstract morphemes and lexical representation: The CV-skeleton in Arabic. *Cognition, 92*(3), 271-303.

Boudelaa, S., & Marslen-Wilson, W. D. (2004b). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language, 90*(1-3), 106-116.

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes, 20*(1-2), 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*(4-6), 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, WD. (2015). Structure, form, and meaning in the mental lexicon: evidence from Arabic. *Language, Cognition and Neuroscience, 30*(8): 955–992.

Bresnan, Joan. (1982). On the Mental Representation of Grammatical Relations. Cambridge, MA: MIT Press.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*(1), 1-15.

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phonemic detection, and learning to read. *Developmental Psychology, 26*, 429-438.

Burani, C., & Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes, 2*(3-4), 217-227.

Burani, C., & Thornton, A. M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 157-207). Berlin: Mouton de Guyter.

Bürki, A., & Gaskell, M. G. (2012). Lexical representation of schwa words: Two mackerels, but only one salami. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 617–631.

Bybee, J. (1995). Diachronic and typological properties of morphology and their implication for representation. In L.B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 225–246). Hillsdale, NJ: Erlbaum.

Cantineau, J. (1950a). La notion de schème et son altération dans diverse langues sémitiques . [The notion of pattern and its modification in different Semitic Languages] *Semitica, III*, 73-83.

Cantineau, J. (1950b). Racines et Schèmes. In M. W (Ed.), *Mélanges* [Roots and patterns] (pp. 119-124). Paris: Maisonneuve et Cie.

Caramazza, A., & Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology/Revue Canadienne De Psychologie, 34*(1), 77-81.

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

Caravolas, M., & Bruck, M. (1993). The effect of oral and written language input on children's phonological awareness: A cross-linguistic study. *Journal of Experimental Child Psychology, 55*(1), 1-30.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*(1), 77-111.

Castles, A., Holmes, V. M., Neath, J., & Kinoshita, S. (2003). How does orthographic knowledge influence performance on phonological awareness tasks? *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 56*(3), 445-467.

Chard, D. J., & Dickson, S. V. (1999). Phonological awareness. instructional and assessment guidelines. *Intervention in School and Clinic, 34*(5), 261-270.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language, 17*(1), 205-215.

Chen, H., & Leung, Y. (1989). Patterns of lexical processing in a non-native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 316-325.

Cheng, Y., & Howard, D. (2008). The time cost of mixed-language processing: An investigation. *International Journal of Bilingualism, 12*(3), 209-222.

Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition: International Journal of Cognitive Science, 102*(3), 341-360.

Cheung, H., & Chen, H. (2004). Early orthographic experience modifies both phonological awareness and on-line speech processing. *Language and Cognitive Processes, 19*(1), 1-28.

Cheung, H., Chen, H., Lai, C. Y., Wong, O. C., & Hills, M. (2001). The development of phonological awareness: Effects of spoken language experience and orthography. *Cognition, 81*(3), 227-241.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 104-119.

Cohen, M. (1951). Langues chamito- sémitiques et linguistique historique. *Scientia, 86*, 304–310.

Cohen-Goldberg, A. M. (2012). Phonological competition within the word: Evidence from the phoneme similarity effect in spoken production. *Journal of Memory and Language, 67*, 184–198.

Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language, 28*(1), 1-13.

Colé, P., & Segui, J. (1994). Grammatical incongruency and vocabulary types. *Memory & Cognition, 22*(4), 387-394.

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.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 1084-1096.

Connine, C. M., Titone, D., Deelman, T. & Blasko, D. (1997) Similarity mapping in spoken word recognition. *Journal of Memory and Language 37*: 463–80.

Cook, V. (2003). Introduction: The changing LI in the L2 user's mind. In V. Cook (Ed.), *Effects of the Second Language on the First (pp. 1-18).* Clevedon: Multilingual Matters.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics*, *9*(1), 1-16.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*(4), 491-511.

Cutler, A. & Norris, D. (1979). Monitoring sentence comprehension. In W.E. Cooper and E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113-134). Hillsdale, NJ: Erlbaum.

Cutler, A., Treiman, R., & van Ooijen, B. (2010). Strategic deployment of orthographic knowledge in phoneme detection. *Language and Speech, 53*(3), 307-320.

Dalrymple-Alford, E. C. (1985). Language switching during bilingual reading. *British Journal of Psychology, 76,* 111–122.

Davis, M. H., van Casteren, M., & Marslen-Wilson, W. D. (2003). Frequency effects in processing inflected Dutch nouns: A distributed connectionist account. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 427-462). Berlin: Mouton de Guyter.

De Cara, B., & Goswami, U. (2003). Phonological neighbourhood density: Effects in a rhyme awareness task in five-year-old children. Journal of Child Language, 30(3), 695–710.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., & Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language, 81*(1-3), 555-567.

De Jong, N. H., Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, *15*(4-5), 329-365.

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.

Diependaele, K., Grainger, J., Sandra D. (2012). Derivational morphology and skilled reading: An empirical overview. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 311-332). Cambridge: Cambridge University Press.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes, 20*(1-2), 75-114.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179-201). New York, NY, US: Oxford University Press, New York, NY.

Dijkstra, T., Moscoso del Prado Martin, F., Schulpen, B., Schreuder, R., & Baayen, R. H. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes, 20*(1-2), 7-41.

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology, 49*(2), 264-71.

Dijkstra, T., & Van Hell, J. G. (2003). Testing the language mode hypothesis using trilinguals. *International Journal of Bilingual Education and Bilingualism, 6*(1), 2-16.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175-224.

Drews, E., & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 1098-1116.

Dufour, S., & Peereman, R. (2003). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition, 88*(3).

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 663-679.

Ehri, I.C. (1975). Word consciousness in readers and pre-readers. *Journal of Educational Psychology, 67(2)*, 204-212.

Eid, M. (1988). Principles for code-switching between standard and Egyptian Arabic. *Al-cArabiyya, 21*(1-2), 51-79.

Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis. *Reading and Writing: An Interdisciplinary Journal, 8*(6), 453-485.

El-Hassan, S. A. (1978a). Educated Spoken Arabic in Egypt and the Levant: A critical review of diglossia and related concepts. *Archivum Linguisticum 8(2),* 112-132.

El-Hassan, S.A. (1978b).Variation in the demonstrative system in Educated Spoken Arabic. *Archivum Linguisticum 11(1),* 32-57.

Elliott, L. L., Hammer, M. A., & Evan, K. E. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception & Psychophysics, 42*(2), 150-157.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V. H. P., Polyzoe, N., Tsipa, M., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39*(4), 438-468.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech, 53*(3), 343-365.

Feldman, L. B., & Andjelkovic, D. (1992). Morphological analysis in word recognition. *Haskins Laboratories Status Report on Speech Research, 111-112(-),* 181-192

Feldman, L. B., & Pastizzo, M. J. (2003). Morphological facilitation: The role of *semantic transparency and family size*. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 233-258). Berlin: Mouton de Guyter.

Ferguson, C. A. (1959). Diglossia. *Word, 15*, 325-340.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51*(1), 1-22.

Ford, M. A., Marslen-Wilson, W. D., & Davis, M. H. (2003). Morphology and frequency: Contrasting methodologies. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 89-124). Berlin: Mouton de Guyter.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680-698.

Foss, D., & Gernsbacher, M. A. (1983). Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning* & *Verbal Behavior,* 22, 609-632.

Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. *Haskins Laboratories Status Report on Speech Research, 105-106*(Jan-June), 53-64.

Frauenfelder, U. H., & Segui, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition, 17*(2), 134-140.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance, 16*(1), 77-91.

Friedman, L., & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, *59*, 127–136.

Frith, U. (1998). Editorial: Literally changing the brain. *Brain*, 121, 1011–1012.

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*(4), 829-856.

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. (2013). *An introduction to statistical learning with applications in R.* London: Springer New York Heidelberg Dordrecht.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*(3), 468-492.

Giraudo, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes, 15*(4-5), 421-444.

Giraudo, H., & Grainger, J. (2003). On the role of derivational affixes in recognizing complex words: Evidence from masked priming. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 89-124). Berlin: Mouton de Guyter.

Glinert, L. (1996). *Modern Hebrew: An essential grammar*. London: Routledge.

Goh, W. D., Suarez, L., Yap, M. J., & Hui Tan, S. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. Psychonomic Bulletin & Review, 16(5), 882-887.

Gonnerman, L. M., Seidenberg, M. S., & Andersen E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, *136*(2), 323-345.

Goswami, U. (1999). The relationship between phonological awareness and orthographic representation in different orthographies. *Learning to read and write: A cross-linguistic perspective* (pp. 134-156). New York, NY, US: Cambridge University Press.

Goswami, U., & Bryant, P. E. (1990). *Phonological skills and learning to read.* Hove, East Sussex, UK: Psychology Press.

Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 39A*(2), 295-319.

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language, 30,* 370–384.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. Journal of Memory and Language, 35, 623 647.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67-104.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267-283.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech.

Language and Cognitive Processes, 3(3), 233-274.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. M. B. de Groot, & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225-254) Mahwah, NJ: Lawrence Erlbaum Associates.

Grosjean, F. (2008). *Studying bilinguals*. Oxford: Oxford University Press.

Guasch, M., Sanchez-Casas, R., Ferre, P., & Garcia-Albea, J. E. (2011). Effects of the degree of meaning similarity on cross-language semantic priming in highly proficient bilinguals. *Journal of Cognitive Psychology, 23*(8), 942-961.

Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology, 4*, 1-13.

Gwilliams, L. & Marantz, A. (2015). Non-linear processing of a linear speech stream: The influence of morphological structure on the recognition of spoken Arabic words. *Brain & Language 147*, 1–13.

Haggan, M. (2007). Text messaging in Kuwait. Is the medium the message? Multilingua, 26, 427–449.

Hallé, P. A., Chereau, C., & Segui, J. (2000). Where is the /b/ in "absurde" [apsyrd]? it is in French listeners' minds. *Journal of Memory and Language, 43*(4), 618-639.

Hay, J., & Baayen, H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences, 9*(7), 342-348.

Heath, J. (2003). Arabic derivational ablaut, processing strategies, and consonantal "roots". In J. Shimron (Ed.*), Language processing and acquisition in languages of Semitic root-based morphology* (pp. 100–115). Amsterdam: John Benjamins.

Heredia, R., & Brown, J. (2004). Code-switching. *In the encyclopedia of linguistics*. Chicago: Fitzroy Dearborn Publishers.

Hilaal, Y. (1990). Deriving from roots and word patterns. *Linguistica Communicatio, 1*, 77-80. (in Arabic).

Hogan, T. P., Bowles, R. P., Catts, H. W., & Storkel, H. L. (2011). The influence of neighborhood density and word frequency on phoneme awareness in 2nd and 4th grades. *Journal of Communication Disorders, 44*(1), 49-58.

Holes, C. (2004). *Modern Arabic; structures, functions, and varieties*. London and New York: Longman.

Holmes, V. M., & O'Regan, J. K. (1992). Reading derivationally affixed French words. *Language and Cognitive Processes, 7*(2), 163-192.

Homer, B. D., & Olson, D. R. (1999). Literacy and children's conception of words. *Written Language and Literacy, 2*(1), 113-140.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29,* 296-305.

Hunt, A. R., & Klein, R. M. (2002). Eliminating the cost of task set reconfiguration. *Memory & Cognition, 30*(4), 529-39.

Ibáñez, A. J., MacIzo, P., & Bajo, M. T. (2010). Language access and language selection in professional translators. *Acta Psychologica, 135*(2), 257-266.

Ibrahim, R. (2006). Do languages with cognate relationships have advantages in second language acquisition? *The Linguistics Journal, 1*(3), 66-96.

Ibrahim, R. (2009). The cognitive basis of diglossia in Arabic: Evidence from a repetition priming study within and between languages. *Psychology Research and Behavior Management, 2*, 93-105.

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

Idrissi, A., Prunet, J., & Beland, R. (2008). On the mental representation of Arabic roots. *Linguistic Inquiry, 39*(2), 221-259.

Isel, F., Baumgaertner, A., Thraen, J., Meisel, J. M., & Buechel, C. (2010). Neural circuitry of the bilingual mental lexicon: Effect of age of second language acquisition. *Brain and Cognition, 72*(2), 169-180.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*, 23–62.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition, 2*(1), 59-75.

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language, 44*(1), 32-51.

Katamba, F., & Stonham, John T. (2006). *Morphology* (2nd ed., Modern linguistics series (Palgrave (Firm))). Basingstoke [England] ; New York: Palgrave Macmillan.

Katz, L., Rexer, K., & Lukatela, G. (1991). The processing of inflected words. *Psychological Research, 53*(1), 25-32.

Kazanina, N. (2011). Decomposition of prefixed words in Russian. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(6),* 1371-1390.

Kiyonaga, K., Grainger, J., Midgley, K. J., & Holcomb, P. J. (2007). Masked cross-modal repetition priming: An event-related potential investigation. Language and Cognitive Processes, 22, 337 376.

Kolers, P. A. (1966). Reading and talking bilingually. *American Journal of Psychology, 79,* 357-376.

Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder, & B. Weltens (Eds.), *The bilingual lexicon* (pp. 53-81). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory (volume 2)* (pp. 389-395). London: Wiley.

Kroll, J. F., & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169-199). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18*(2), 137-171.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes, 23,* 1089–1132.

Kurvers, J., & Uri, H. (2006). Metalexical awareness: Development, methodology or written language? A cross-linguistic comparison. *Journal of Psycholinguistic Research, 35(4),* 353-367.

Kuznetsova, A., Brockhoff, P. & Christensen, R. (2016). lmerTest: Tests in linear mixed effect models. Version 2.0-25. Retrieved from: *http://cran.r project.org/web/packages/lmerTest/lmerTest.pdf*

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, *16*(3), 508-517.

Laks, L., Berman, R. A. (2014). A new look at diglossia: Modality-driven distinctions between spoken and written narratives in Jordanian Arabic. In Elinor Saiegh-Haddad & R. Malatesha Joshi (Eds.), *Arabic literacy: Theoretical insights and practical challenges* (pp. 241-254). Dordrecht: Springer.

Levin, I., Saiegh-Haddad, E., Hende, N., & Ziv, M. (2008). Early literacy in Arabic: An intervention study among Israeli Palestinian kindergartners. *Applied Psycholinguistics, 29*, 413-436.

Léwy, N. & Grosjean, F. (in progress). The computerized version of BIMOLA: A bilingual model of lexical access. Manuscript, University of Neuchâtel, Switzerland.

Li, P. (1996). Spoken word recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory and Language*, *35(6)*, 757-774.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Liberman, I. Y. (1971). Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society, 21*, 71-87.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*(2), 201-212.

Lieber, Rochelle 1992. Deconstructing Morphology. Chicago: University of Chicago Press.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In Ferguson, Charles A., Menn, Lise, & Stoel-Gammon, Carol (Eds.), *Phonological development: Models, research, implications* (pp. 131-163). Timonium, MD: York Press, Inc.

Liu, P.L. (2014). Using Eye Tracking to Understand the Responses of Learners to Vocabulary Learning Strategy Instruction and Use. *Computer Assisted Language Learning, 27*(4), 330-343.

Luce, P., & Pisoni, D. B. (1998). Recognizing spoken words. The neighborhood activation model. Ear & Hearing, 19, 1-36.

Lüdeling, A. & de Jong, N. H. (2002). German particle verbs and word formation. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Explorations in verb- particle constructions* (pp. 315-333). Berlin: Mouton de Gruyter.

Maamouri, M. (1998a). Arabic diglossia and its impact on the quality of education in the Arab region, *A paper presented at The World Bank Mediterranean Development Forum, Marrakech*, Morocco, September, 3-6, 1998.

Maamouri, M. (1998b). *Language education and human development: Arabic diglossia and its impact on the quality of education in the Arab region.*

Macizo, P., Bajo, T., & Paolieri, D. (2012). Language switching and language competition. *Second Language Research, 28(2),* 131-149.

Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behaviour, 10,* 480–487.

Mahfoudhi, A. (2005). *Morphological and phonological units in the Arabic mental lexicon: Implications for theories of morphology and lexical processing* (Ph.D.). Available from ProQuest Dissertations & Theses A&I. (305358388).

Mahfoudhi, A. (2007). Roots and patterns in Arabic lexical processing. In E. Benmamoun (Ed.), *Perspectives on Arabic linguistics, XIX* (pp. 123-150) Benjamins.

Manelis, L., & Tharp, D. A. (1977). The processing of affixed words. *Memory & Cognition, 5*(6), 690-695.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71-102.

Marslen-Wilson, W. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*(1), 3-33.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101***,** 3–33.

Marslen-Wilson, W., & Zhou, X.-L. (1999). Abstractness, allomorphy, and lexical architecture. *Language and Cognitive Processes, 14,* 321–352.

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh, & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 133-147). MIT Press.

McBride-Chang, C., Bialystok, E., Chong, K. K. Y., & Li, Y. (2004). Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology, 89*(2), 93-111.

McCarthy, J. J. (1979). Formal problems in Semitic phonology and morphology. Unpublished Ph.D. Dissertation, MIT.

McCarthy, J. J. (1981). A prosodic theory of non-concatenative morphology. *Linguistic Inquiry, 12*, 373-418.

McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

Meiseles, G. (1980). Educated Spoken Arabic and the Arabic Language Continuum. *Archivum Linguisticum, 11*(2):118-148.

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition, 25*(1), 47-56.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*(1), 3-19.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. *Word recognition in beginning literacy* (pp. 89-120). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

Meunier, F., & Segui, J. (1999). Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language, 41*(3), 327-344.

Meunier, F., & Segui, J. (2002). Cross-modal morphological priming in French. *Brain and Language, 81*(1-3), 89-102.

Meunier, F., & Segui, J. (2003). Identification of spoken prefixed words in French. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 337-353). Berlin: Mouton de Guyter.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25-40.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234.

Mimouni, Z., Kehayia, E., & Jarema, G. (1998). The mental representation of singular and plural nouns in Algerian Arabic as revealed through auditory priming in agrammatic aphasic patients. *Brain and Language, 61*(1), 63-87.

Monsell, S. (2003). Task switching. *Trends Cogn Sci., 7*(3), 134-140.

Morais, J. (1991). Constraints on the development of phonemic awareness. In Liberman, Brady, Susan A., & Shankweiler, Donald P. (Eds.), *Phonological processes in literacy: A tribute to Isabelley* (pp. 5-27). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*(1-2), 45-64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*(4), 323-331.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An eletrophysiological study of code switching. *Brain and Language, 80*(2),188-207.

Moscoso del Prado Martin, F., Bertram, R., Haikio, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1271-1278.

Moscoso del Prado Martin, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language, 53*(4), 496-512.

Moscoso del Prado Martin, F., Kostic, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition, 94*(1), 1-18.

Mostari, H. A. (2009). What do mobiles speak in Algeria? Evidence from language. Current Issues in Language Planning, 10, 377–386.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology, 102*(6), 963-968.

Muysken, P. (2000). *Speech: A typology of code-mixing*. Cambridge, UK: Cambridge University Press.

Neijt, A., Schreuder, R., & Baayen, H. (2003). Verpleegsters, ambassadrices and masseuses: Stratum differences in the comprehension of Dutch words with feminine agent suffixes. *Linguistics in the Netherlands, 20*, 117-127.

Newman, R. S., Sawusch, J. R. & Luce, P. A. (1997) Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and*

Performance 23(3): 873–89.

Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, *15*(4-5), 389-420.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences,23*(3), 299-370.

Olson, D. R. (1996). Towards a psychology of literacy: On the relations between speech and writing. *Cognition, 60*(1), 83-104.

Owens, J. (2001). Arabic Sociolinguistics. *Arabica, 48*(4), 419-469.

Pavlenko, A. (2000). L2 Influence on LI in late bilingualism..Issues in Applied *Linguistics, 11*, 175-205.

Pavlenko, A. (2003). 'I feel clumsy speaking Russian': L2 influence on LI in narratives of Russian L2 users of English. In V. Cook (Ed.), *Effects of the Second Language on the First* (pp. 32-61). Clevedon: Multilingual Matters.

Pavlenko, A. (2004). L2 influence and LI attrition in adult bilingualism. In M. S. Schmid, B. Kopke, M. Keijzer, & L. Weilemar (Eds.), *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* (pp. 47-59). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Perea, M., Dunabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language, 58*(4), 916-930.

Petersson, K., Reis, A., Askelöf, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. *Journal of Cognitive Science, 12*(3), 364–382.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes, 15*(4-5), 445-485.

Proverbio, A. M., Leoni, G., & Zani, A. (2004). Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia, 42*(12), 1636-1656.

Prunet, J., Beland, R., & Idrissi, A. (2000). The mental representation of Semitic words. *Linguistic Inquiry, 31*(4), 609-648.

Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition, 91*(3), B35-B45.

Qasem, M., & Foote, R. (2010). Crosslanguage lexical activation: A test of the revised hierarchical and morphological decomposition models in arabic-english bilinguals. *Studies in Second Language Acquisition, 32(1),* 111-140.

Radeau, M., Morais, J., Mousty, P., & Bertelson, P. (2000). The effect of speaking rate on the role of the uniqueness point in spoken word recognition. Journal of Memory and Language, 42, 406 -422.

Radeau, M., Mousty, P., & Bertelson, P. (1989). The effect of the uniqueness point in spoken-word recognition. *Psychological Research*, *51*, 123–128.

Rajhi, A. (1976). *'Attatbiiq 'assarfi* [Practical Morphology]. Lebanon: Dar Annahdha Al-Arabiyya.

Ranbom, L. J., & Connine, C. M. (2011). Silent letters are activated in spoken word recognition. Language and Cognitive Processes, 26(2), 236-261.

Ratcliffe, R. R. (1998). *The "broken" plural problem in Arabic and comparative Semitic: Allomorphy and analogy in non-concatenative morphology* Amsterdam: John Benjamins.

Ratcliffe, R. R. (2004). Sonority-based parsing at the margins of Arabic morphology: In response to Prunet, Beland, and Idrissi (2000) and Davis and Zawaydeh (1999, 2001). *Al-ʿArabiyya: Journal of the American Association of Teachers of Arabic*, 37, 53-75.

Ravid, D., & Malenky, A. (2001). Awareness of linear and nonlinear morphology in Hebrew: A developmental study. *First Language, 21*(1), 25-56.

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*(1-2), 31-44.

Reid, A. A., & Marslen-Wilson, W. D. (2003). Lexical representation of morphologically complex words: Evidence from Polish. In R.H. Baayen, & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 287-336). Berlin: Mouton de Guyter.

Roth, F. P., Troia, G. A., Worthington, C. K., & Handy, D. (2006). Promoting awareness of sounds in speech (PASS): The effects of intervention and stimulus characteristics on the blending performance of preschool children with communication impairments. Learning Disability Quarterly, 29, 67–88.

Russak, S., & Saiegh-Haddad, E. (2011). Phonological awareness in Hebrew (L1) and English (L2) in normal and disabled readers. Reading and Writing, 24(4), 427-442.

Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition, 17*(3), 542-555.

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

Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudowords in a diglossic context. *Applied Psycholinguistics, 25*(4), 495-512.

Saiegh-Haddad, E. (2007a). Epilinguistic and metalinguistic phonological awareness may be subject to different constraints: Evidence from Hebrew. *First Language, 27*(4), 385-405.

Saiegh-Haddad, E. (2007b). Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics, 28*(4), 607-625.

Saiegh-Haddad, E. & Henkin-Roitfarb, R. (2014). The Structure of Arabic Language and Orthography. In E. Saiegh-Haddad, & R. Joshi (Eds), *Handbook of Arabic Literacy: Insights and Perspectives* (Vol. 9, Literacy Studies) (pp. 3-28). Dordrecht: Springer Netherlands.

Saiegh-Haddad, E., Kogan, N., & Walters, J. (2010). Universal and language-specific constraints on phonemic awareness: Evidence from Russian-Hebrew bilingual children. *Reading and Writing, 23*(3-4), 359-384.

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.

Sanchez-Casas, R., Ferre, P., Garcia-Albea, J. E., & Guasch, M. (2006). The nature of semantic priming: Effects of the degree of semantic similarity between primes and targets in spanish. *The European Journal of Cognitive Psychology, 18*(2), 161-184.

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition, 37*(5), 569-586.

Schoonbaert, S., Holcomb, P. J., Grainger, J., & Hartsuiker, R. J. (2011). Testing asymmetries in noncognate translation priming: Evidence from RTs and ERPs. *Psychophysiology, 48*(1), 74-81.

Schreier, D. (1998). Recovery time during the processing of code-switches in bilinguals. *Travaux Neuchatelois De Linguistique (TRANEL), 28*(May), 103-123.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. *Morphological aspects of language processing* (pp. 131-154). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language, 37*(1), 118-139.

Schreuder, R., Burani, C., & Baayen, R. H. (2003). Parsing and semantic opacity. In E. M. H. Assink, & D. Sandra (Ed.), *Neuropsychology and cognition: reading complex words: Cross-language studies* (pp. 159-190). Kluwer Academic Publishers.

Schriefers, H., Zwitserlood, P., & Roelofs, A. (1991). The identification of morphologically complex spoken words: Continuous processing or decomposition? *Journal of Memory and Language, 30,* 26–47.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*(2), 197-212.

Schwarz, G. E. (1978). Estimating the dimension of a model, *Annals of Statistics 6 (2): 461–464.*

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition, 25*(4), 425-437.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143-174.

Share, D. L., & Blum, P. (2005). Syllable splitting in literate and preliterate Hebrew speakers: Onsets and rimes or bodies and codas? *Journal of Experimental Child Psychology, 92*, 182-202.

Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect of lexical access. *Memory & Cognition, 12*(4), 380-386.

Stanners, R. F., Neiser, J. J., & Painton, S. (1979). Memory representation for prefixed words. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 733-743.

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon, 1*(1), 85-123.

Studdert-Kennedy, M. (1984). Sources of variability in early speech development. *Haskins Laboratories Status Report on Speech Research, 77-78*(Jan-June), 1-21.

Suleiman, S. M. (1985). Jordanian Arabic between diglossia and bilingualism: Linguistic analysis. *Pragmatics and Beyond, 6*(8), 1-131.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education Inc.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7*(4), 263-272.

Taft, M. (1994). Interactive activation as a framework for understanding morphological processing. *Language and Cognitive Processes 9***,** 271-294.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology, 57A*, 745–765.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58*(2), 366-379.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 638.

Taft, M., Hambly, G., & Kinoshita, S. (1986).Visual and auditory recognition of prefixed words. *Quarterly Journal of Experimental Psychology,* 38A, 357-366.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2*(1), 45-58.

The Jordan Academy of Arabic. (2006). *muEjзam 'alfaaD 'aHayaah lEaamyah fi l'urdun [Dictionary of the Everyday Language in Jordan]*. Lebanon: Lebanon Bookshop Publisher.

Thomas, M. S. C., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language, 43*(1), 44-66.

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1-L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1412-1431.

Traficante, D., & Burani, C. (2003). Visual processing of Italian verbs and adjectives: The role of the inflectional family size. In R.H. Baayen, & R. Schreuder (Eds), *Morphological Structure in Language Processing* (pp. 287-336). Germany, Berlin: Mouton de Guyter.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*(1-3), 49-74.

Treiman, R. (1988). The internal structure of the syllable. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 27-52). Norwell, MA: Kluwer Academic.

Treiman, R., & Zukowski, A. (1991). Levels of phonological awareness. In Liberman, Brady, Susan A., & Shankweiler, Donald P. (Ed.), *Phonological processes in literacy: A tribute to Isabelley* (pp. 67-83). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Troia, G. A., Roth, F. P., & Yeni-Komshian, G. H. (1996). Word frequency and age effects in normally developing children's phonological processing. *Journal of Speech and Hearing Research, 39*(5), 1099-1108.

Tyler, L. K., Marslen-Wilson, W., Rentoul, J., & Hanney, P. (1988). Continuous and discontinuous access in spoken word-recognition: The role of derivational prefixes. *Journal of Memory and Language*, 27(4), 368-381.

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling the stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 760-771.

Ussishkin, A. (1999). The inadequacy of the consonantal root: Modern hebrew denominal verbs and output-output correspondence. *Phonology, 16*(3), 401-442.

Ussishkin, A. (2005). A fixed prosodic theory of nonconcatenative templatic morphology. Natural Language & Linguistic Theory, 23(1), 169-218.

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*(3), 193-211.

Ventura, P., Kolinsky, R., Fernandes, S., Querido, L., & Morais, J. (2007). Lexical restructuring in the absence of literacy. Cognition, 105, 334–361.

Versteegh, K. (1997). *The Arabic language*. Edinburgh, UK: Edinburgh University Press.

von Studnitz, R. E., & Green, D. W. (1997). Lexical decision and language switching. *International Journal of Bilingualism, 1*(1), 3-24.

von Studnitz, R. E., & Green, D. W. (2002). The cost of switching language in a semantic categorization task roswitha E. von studnitz and david W. green the cost of switching language. *Bilingualism, 5*(3), 241-251.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779-804.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13*(3), 286-350.

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing, 16*(1-2), 5-20.

Walley, A. C., Michela, V. L., & Wood, D. R. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Perception & Psychophysics, 57*(3), 343-351.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1-25.

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. Journal of Memory and Language, 68(4), 362-378.

Wimmer, H., Landerl, K., & Schneider, W. (1994). The role of rhyme awareness in learning to read a regular orthography. *British Journal of Developmental Psychology, 12*(4), 469-484.

Wright, W. (1967). *A grammar of the Arabic language*. Cambridge: Cambridge University Press.

Wurm, L. H. (1997). Auditory processing of prefixed English words is both continuous and decompositional. *Journal of Memory and Language, 37*(3), 438-461.

Wurm, L. H., & Aycock, J. (2003). Recognition of spoken prefixed words: The role of early conditional root uniqueness points. In R. H. Baayen, & R. Schreuder (Eds), *Morphological Structure in Language Processing* (pp. 259-286). Berlin: Mouton de Guyter.

Wurm, L. H., Ernestus, M. T. C., Schreuder, R., & Baayen, R. H. (2006). Dynamics of the auditory comprehension of prefixed words: Cohort entropies and conditional root uniqueness points. *The Mental Lexicon, 1*(1), 125-146.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language, 72,* 37-48.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683-689.

Ziegler, J. C., & Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: A developmental study. *Psychonomic Bulletin & Review, 14*(1), 75-80.