Decomposability and the effects of morpheme frequency in lexical access

by

Samantha Carol Wray

A Dissertation Submitted to the Faculty of the

Department of Linguistics

In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

In the Graduate College

The University of Arizona

2 0 1 6

The University of Arizona
Graduate College

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Samantha Wray, titled Decomposability and the effects of morpheme frequency in lexical access and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

____ Date: 20 July 2016
Adam Ussishkin

____ Date: 20 July 2016
Michael Hammond

____ Date: 20 July 2016
Samira Farwaneh

Final approval and acceptance of this dissertation is contingent upon the candidates submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 20 July 2016
Dissertation Director: Adam Ussishkin

Statement by Author

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

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

Signed: Samantha Wray

Acknowledgments

I am incredibly grateful for all of the support and help I received throughout the completion of this dissertation. I am deeply honored by the guidance and mentoring provided to me by my ever-reliable advisor Adam Ussishkin, and thank him for his invaluable wisdom and remarkable support. I also extend my deepest gratitude to the rest of my committee, Michael Hammond and Samira Farwaneh, for their incredible expertise and kindness. I am very grateful to Andrew Wedel, Amy Fountain, Natasha Warner, and Sandiway Fong for the training and support they provided.

I would like to thank my colleagues at the University of Arizona for their immeasurable support and fellowship, especially the incomparable Rolando Coto-Solano, as well as Dane Bell, Julia Fisher, Andréa Davis, Amy LaCross, Kevin Schluter, Deniz Tat, Lionel Mathieu, Alex Trueman, Bryan James Gordon, and Keri Miller. I'm also very thankful to Wyatt Welch and Stone Runyan for their enthusiasm and kindness.

I am very grateful for the assistance from my colleagues at QCRI: Stephan Vogel, Ahmed Ali, Hamdy Mubarak, Yasser Makram, and Abdurahman Ghanem, as well as my colleagues at the University of Maryland: Tim Buckwalter, Valerie Novak, C. Anton Rytting, Jennifer Boutz, Mona Madgavkar, and David Zajic. Thank you all so very much for imparting your expertise and for being so helpful.

In addition, I would like to give sincere thanks for the educators at Western Washington University, including the wonderful mentorship provided by Rudolf Weiss, Kristin Denham, Anne Lobeck, Shaw Gynan, Todd Haskell, and Dwan Shipley; and my WWU colleagues Marten van Schijndel, Lauren Apfelbaum, and Andrew Blick. Thank you so much for inspiring me and fostering my love for linguistics.

I am indebted to the assistance received onsite in Amman, first and foremost to Rasha Jaloukh. I also wish to thank Qasid Institute for space and assistance from many dedicated educators, including Khalid Snobar, Banan Amar, Manal Odeh, and Omar Matadar. Thank you also to the University of Jordan for space and assistance from the staff including Rula Qawwas and Manal Yusifi. Thanks are also due to Hassan Shiban and Shahreena Shahrani for their input.

For their years of love and support, I would like to thank my very beloved mother, father, and brothers. Thank you all so much for your help. Thanks also to Holly for being such a good friend.

Finally, I would like to express my gratitude for the funding for this dissertation received from the National Science Foundation under BCS-1533780, the University of Arizona Graduate and Professional Student Council, and the University of Arizona Graduate Interdisciplinary Programs Raphael and Jolene Gruener Research Travel Award. An allocation of computer time from the UA Research Computing High Performance Computing (HPC) and High Throughput Computing (HTC) at the University of Arizona is gratefully acknowledged.

Table of Contents

Table of Contents—*Continued*

Table of Contents—*Continued*

List of Figures

List of Tables

Abstract

This dissertation addresses an unanswered question in Arabic psycholinguistics. Arabic words are characterized by their nonconcatenative structure, in which a consonantal root that encodes the main semantic content is interleaved with a derivational pattern (called "binyan", pl. "binyanim"), which is typically vocalic but may also contain consonantal elements and contributes grammatical information. The canonical example of the Semitic root and binyan system is the combination of root /ktb/ which denotes the broad semantic sense of 'writing' with verbal binyan /CaCaC/ (with C indicating a root consonant) to form [katab] 'he wrote' and with nominal place binyan /maCCaC/ to form [maktab] 'office'. Although significant work has been done on the psycholinguistic reality of Arabic morphemes by exploring various phonological, morphological and semantic features across numerous experimental modalities in both the visual and auditory domains (Boudelaa and Marslen-Wilson, 2004b, 2005, 2011), no study has investigated the roles of base/morpheme frequency and surface/word frequency and their implications for underlying morphological structure in the lexicon of Arabic as has been done for English, Dutch, and Finnish (Baayen et al., 1997; Alegre and Gordon, 1999; New et al., 2004; Taft, 1979, 2004).

Competing models of word recognition propose various integrations of morphology. Whole-word models suggest that there are no separate representations for morphemes, and that the co-activation of related words can be attributed to similarity in form and meaning (Norris and McQueen, 2008; Tyler et al., 1988). Decomposition models posit that words are recognized by accessing the words' constituent morphemes (Meunier and Segui, 1999; Taft et al., 1986; Wurm, 2000). Hybrid models incorporate multiple pathways to recognition. Words are either recognized holistically or by their constituent morphemes depending on multiple factors (Balling and Baayen, 2008; Taft and Nguyen-Hoan, 2010; López-Villaseñor, 2012). Of most relevance to the current

study is the role of the productivity of the words' derivational affixes: words with unproductive affixes are processed holistically whereas words with productive affixes are processed as a function of their morphemes.

This dissertation presents results from four auditory lexical decision experiments performed with native Jordanian speakers in Amman, Jordan, and provides evidence that binyan productivity determines whether the frequency of the base morpheme affects the speed of recognition. By manipulating root and word frequency for three binyanim, one more productive and two less productive, I provide evidence that verbs in the productive binyan are fully decomposable during lexical access and verbs in less productive binyanim are recognized holistically. For a more productive binyan, I examine Binyan I of the form /CaCaC/, and two less productive binyanim are Binyan VIII of the form /iCtaCaC/and Binyan X of the form /staCCaC/.

These results together support a hybrid model of lexical access in which some words are recognized via decomposition into the morphemes they are composed of, and others are recognized by their whole word form. These results are consistent with those of Balling and Baayen (2008); Taft and Nguyen-Hoan (2010); Bertram et al. (2000b), among others, as derivational affix productivity is the deciding factor determining whether a word will be recognized holistically or decomposed during lexical access.

1. Arabic verbs in the mental lexicon

A key feature of the study of lexical access is the extent to which the morphological structure of a word affects the speed of word recognition, as well as implications for the structure of the mental lexicon and the mechanisms utilized during lexical access. Much evidence has shown that morphology plays a role in lexical access, and operates independently from processing gains contributed by semantic overlap in both the auditory and visual domain, phonology in the auditory domain (Grainger et al., 2005; Yates et al., 2004), and orthography in the visual domain (Smolka et al., 2009). A great deal of research concerned with the role of morphological structure and the decomposibility of words into their constituent morphemes during lexical access has been centered around languages which typically utilize linear morphology. The focus of much psycholinguistic research aiming to elucidate the nature of the lexical access process has been English and other Indo-European languages, which largely use a word formation system entailing stems that are prefixed and suffixed in a concatenative manner. For example, in a word such as *reintegrated*, the morphemes *re-* and *-ed* are affixed to the beginning and end of the stem *integrate*. Various competing models proposing different incorporations of morphemes into word recognition processes have been proposed.

Although significant work (outlined in detail below) has been done on the issues of morphological structure and lexical access in Semitic languages, additional aspects of the role of morphology during processing remain to be explored. Relevant to the current study is the relationship between morpheme frequency and word frequency during lexical access in the auditory domain and the implications of derivational verbal affix productivity during this process.

In this chapter, I provide an overview the relevant research in word recognition and consider the various models proposed to illustrate effects of morphological structure in word processing. I also introduce Arabic morphology and discuss the psycholinguistic implications of Arabic verb structure. Finally, I present the focus of the dissertation as an investigation of the roles of root and word frequency in auditory word recognition in Arabic, and I provide an outline of the contents of the following chapters of the dissertation.

1.1. Lexical access and the mental lexicon

There are a number of experimental tasks which have traditionally been used to investigate the dynamics of lexical access and have been used to measure the speed of the word recognition process. Of greatest relevance to the current proposal is the lexical decision task (developed by Meyer and Schvaneveldt (1971), among others), in which a participant is presented a stimulus and asked to decide if it is a real word of their language or not. The lexical decision task is especially sensitive to frequency effects in comparison to other experimental tasks such as naming, in which a participant is presented with a non-linguistic indication of an object or concept and they must name it, and categorization, in which a participant must cluster items together (Balota and Chumbley, 1984). Also, different word features are more salient in a lexical decision task. For example, it has been observed that lexical decision is influenced by more factors, such as semantics, than naming (Baayen et al., 2007).

The speed at which a listener recognizes a word, as measured by the lexical decision task, is determined by a variety of factors, and these can be tested by performing the task in either the visual or auditory domain. The first determines the speed of recognition during reading, and the second determines recognition speed while listening. A cross-modal task would involve stimuli presented both visually and auditorily. The lexical decision task can be constructed so that a participant responds to a target word with no further stimuli, or such that a participant responds to a target word which has been preceded by a secondary stimulus known as a prime. The second task is performed to determine the effect one word has on another, if any. For example, it has been shown that readers respond more quickly to words within one semantic domain if they have just been exposed to a word in the same domain as opposed to an unrelated word (Meyer and Schvaneveldt, 1971). This suggests that semantic activation plays a role in the word recognition process. Priming tasks may be performed in such a way that the participant is aware of both the prime and the target, and can consciously perceive both of them, but is directed to respond to only the latter. Masked priming tasks, in contrast, are constructed such that the prime is perceivable subconsciously but the participant is not aware they have seen or heard it. This can be accomplished by presenting the stimulus for a very short time (such as 50 milliseconds) and sandwiching the presentation of the stimulus between two non-linguistic presentations in the same domain, such as a line of hash marks in the visual domain, or speech played backwards and attenuated by volume in the auditory domain.

Throughout these experimental paradigms, many effects of factors which influence speed of recognition have been attested. First, more frequent words are responded to faster than less frequent words (Luce, 1986; Marslen-Wilson, 1990; Slowiaczek and Pisoni, 1986; Taft et al., 1986). Also, for auditory priming tasks, prime-target pairs facilitate lexical access in cases of shared semantics (Radeau, 1983), shared morphemic relationship (Emmorey, 1989; Marslen-Wilson et al., 1994; Kempley and Morton, 1982), as well as two attested effects for phonological overlap in prime-target pairs, facilitation (Slowiaczek and Hamburger, 1992; Radeau et al., 1995), and inhibition (Slowiaczek and Hamburger, 1992; Radeau et al., 1989; Goldinger et al., 1992). Although many of these effects are also observed in the visual domain, visual and auditory processing diverge in non-trivial ways. For example, although the effects of phonology have been shown to have a potential to inhibit lexical access in the auditory domain, shared phonological relationship between prime-target pairs in a visual lexical decision task results in facilitation (Lukatela and Turvey, 1994).

The contributions of these factors, as well as the differences observed between the visual and auditory modalities, are accounted for in numerous models elucidating the mechanisms that drive lexical access and the resulting effects. Many models of psycholinguistics utilize the concept of the mental lexicon, the collection of words stored in our cognitive system. To be identified, a word must be distinguished from other words in the lexicon. For words which are multimorphemic, the mental lexicon has several potential forms. Are complex words accessed via separate morphemes, full forms, or both? The competing theoretical models aiming to explain morphological effects during lexical access can be broadly categorized into three types:

- models which posit that words are recognized by their complete surface form
- models involving the automatic decomposition of a complex word into its constituent morphemes
- hybrid models which incorporate multiple decomposition-related pathways for recognition

Evidence from non-Semitic languages for each of these model types are discussed in the following subsections. Semitic languages are considered in Section 1.3.

1.1.1. Whole-word storage

Whole-word models posit that there are no separate representations for morphemes, but that related words might co-activate due to similarity in form and meaning (Tyler et al., 1988; Rueckl et al., 1997; Joanisse and Seidenberg, 1999; Seidenberg and Gonnerman, 2000). Under such a model, the word ‘agreement’ would be processed as a function of the frequency of ‘agreement’ without an effect of the frequency of either ‘agree’ or ‘-ment,’ but a priming experiment in which the prime-target pair shared some morphological relation may result in priming because of meaning and form similarities. For example, words may be encoded in the lexicon based on patterns of semantics and phonology, and similar words are represented by similar patterns of these encodings. Thus, words such as 'cat' and 'cap' are encoded by a similar pattern of phonology and therefore co-activate due to this similar pattern, and differ on their semantic patterning (Gonnerman et al., 2007; Plaut and Gonnerman, 2000).

One such model involving this architecture is the TRACE model (McClelland and Elman, 1986a), and its successor TRACE II (McClelland and Elman, 1986b), which utilizes interactions between the phoneme and word level. During perception in the TRACE model, when the auditory signal is perceived, words which share the same phonemes are activated. Because the auditory signal unfolds in time, as each phoneme is heard, the activation pattern changes. Thus, words with overlapping onsets would be potential competitors and be activated as the listener hears them.

The Neighborhood Activation model (Luce and Pisoni, 1998) also incorporates sound-word interaction at the lexical activation level, positing that a perceived signal is compared to stored acoustic-phonetic representations. Furthermore, stored word frequency information is activated to discriminate between competing lexical candidates.

The Cohort model (Marslen-Wilson and Welsh, 1978) is characterized by the 'uniqueness point'. This is the point in the auditory signal in which competitors during the lexical access process become ruled out. Modifications to the model do incorporate separate morphemic storage for prefixes, but do not list all morphemes separately (Tyler et al., 1988; Tyler and Marslen-Wilson, 1986).

Additional whole-storage models include Shortlist A (Norris, 1994) and Shortlist B (Norris and McQueen, 2008). These manifestations of Shortlist differ on their incorporation of likelihood and probability, allowing the newer version Shortlist B to account for frequency effects. Shortlist involves two stages of lexical access. The first is bottom-up activation of lexical candidates and the second is competition between the activated lexical candidates.

Although these models are consistent with effects observed for languages with largely linear morphology such as English and Dutch, their predictions don't hold for Semitic languages since these languages are characterized by their nonconcatenative morphology. This means that in contrast with concatenatively morphological languages such as English and Dutch, the constituent morphemes of a complex word overlap during their unfolding in time with the acoustic signal. Specifically, as English and Dutch rely on prefixation and suffixation for word formation, the idea of a uniqueness point is consistent with morphemes unfolding linearly in time with the acoustic signal. However, given that Arabic words are formed via the interleaving of noncontiguous morphemes (outlined in detail below), these models are not compatible with an Arabic word's internal structure.

1.1.2. Decomposition into morphemes

In contrast, decomposition models entail a process in which complex words are accessed by way of their constituent morphemes. There is a great deal of research on morpheme processing during lexical access showing that multimorphemic words are decomposed into their morphemes and that they share representation in the mental lexicon (Kempley and Morton, 1982; Meunier and Segui, 1999; Marslen-Wilson et al., 1994; Bradley, 1980).

Ample evidence for a specific role of morphological structure during lexical access comes from the fact that inflectional and derivational morphemes are processed differently. For example, Tyler and Cobb (1987) presented evidence from an aphasic patient who could process derivational morphemes, but not inflectional morphemes. Tyler and Cobb (1987) performed an experiment in which an aphasic patient had to monitor for a specific word in a speech string and press a button to indicate when he had heard the word. For each sentence, the monitored word was immediately preceded by a morphologically complex test word that contained either a derivational or inflectional affix. The word was either of a form appropriate for the context, for example "He was the most *wasteful* cook", or was inappropriate for the context such as in the sentence "He was the most *wastage* cook". A control group exhibited delayed responses when the inappropriate affixes were utilized for both inflectional and derivational affixes. However, when compared to a control group, the aphasic speaker exhibited additional delay for inflectional items which were appropriately affixed, and no additional delay for derivational items which were appropriately affixed. This suggests that inflectional and derivational affixes are processed differently.

Models which posit that all words are decomposed into their constituent morphemes during lexical access can be described as containing obligatory decomposition. Evidence for obligatory decomposition is found in Marslen-Wilson et al. (1997) in which they studied processing of past conjugations in English in patients with left hemisphere damage. They performed an auditory priming task and found normal priming for irregular past-tense pairs like 'find/found', but significantly reduced priming for regular past-tense pairs such as 'join/joined'. They explain these effects by appealing to a situation in which irregular forms must be learned and stored whole-form, whereas regular conjugations are composed, and therefore decomposed during recognition.

Indeed, there is evidence that a complex word is decomposed into its morphemes even in the absence of a true morphological relationship early in the lexical access process. In an auditory same-different task in which participants had to listen to words spoken by two different talkers and decide if they had heard the same word twice, Tyler et al. (2002) found that patients had impaired performance not only with regularly inflected word pairs such as 'play/played', but also with nonwords with a pseudomorphemic '-ed' which could be misanalyzed as a past form. Thus, patients also exhibited impaired performance at the same-different task when presented with nonwords such as 'snay/snade'. They conclude that the lexical access system must be able to isolate and identify individual morphemes, and nonwords which conform to a characteristic morpho-phonemic pattern are therefore potentionally decomposable.

1.1.3. Hybrid models

Finally, hybrid models involve a lexical architecture in which words can be represented as both separate morphemes and whole words (Burani and Caramazza, 1987; Schriefers et al., 1992; Balling and Baayen, 2008), predicting a robust effect of the frequency of ‘agreement’ as well as its constituent morphemes. In these models, both the frequency of the morphemes ‘agree’ and ‘-ment’ have an effect on processing during lexical access of the word ‘agreement’. Whereas a simple word exhibits a robust influence of word frequency in reaction time (henceforth RT) facilitation to lexical decision tasks, RTs to a complex word can be affected not only by the its own frequency (surface frequency) but also by the frequency of the morphemes it consists of (base frequency) (Domínguez et al., 2000; Meunier and Segui, 1999). The base frequency of a word like ‘agreement’ is calculated as the cumulative frequency of all of the words which contain the base morpheme ‘agree’, such as ‘agreed’, ‘disagree’, ‘agrees’, and so forth. A base frequency effect occurs when words with high frequency bases are responded to more quickly than words with low frequency bases. In addition, when base frequency is constant, words with a higher surface frequency will be recognized faster (Taft, 1979; Burani et al., 1984). Finally, for regularly inflected words (Baayen et al., 1997; Burani et al., 1984; Schriefers et al., 1992) when whole word frequency is constant, base frequency facilitates lexical decision, and vice versa.

Much of the work related to the relationship between multimorphemic words in the lexicon and frequency has been focused on inflectional morphology. However, Ford et al. (2010) investigated derivational morphology and the implications for lexical access, arriving at an analysis which supports a hybrid model of lexical access. Considering that there are more unproductive derivational affixes than inflectional affixes in English, Ford et al. (2010) investigated the relationship between the number of words which contain an affix and the frequency of these words. They appeal to the concept of family size, which is the number of unique words which contain a given morpheme (Schreuder and Baayen, 1997). For example, the word ‘calculate’ belongs to a morphological family with nine members including itself, and ‘miscalculate’, ‘calculable’, ‘calculation’, ‘calculator’, ‘calculus’, ‘incalculable’, ‘incalculably’, and ‘miscalculation’ (De Jong et al., 2000). This is contrasted with the base frequency of a morpheme, which is the cumulative frequency of all of these words, not a type count of the words themselves. They investigated the roles of affix productivity as well as affix family size separately, and discovered no interaction, indicating the two factors independently influenced RTs. They presented evidence that morpheme frequency depended on suffix productivity, as evidenced by the fact that base morpheme frequency facilitated RTs to productively suffixed words. Family size effects were different - family size facilitated responses independent of the productivity. Thus, the more unique words with the affix (i.e. the bigger the morphological family), the faster response time.

One possible explanation for having separate lexical listings for the surface word and its morphemes is the frequency threshold. For regularly inflected English words, Alegre and Gordon (1999); Gordon and Alegre (1999) proposed that a threshold of 6 occurrences per million resulted in separate lexical representatives of a word and its morphemes, and words below that threshold did not. They determined this threshold based on the results of a series of lexical decision experiments which showed that the frequency of morphologically complex words did not facilitate lexical access speed for words below the threshold, whereas higher frequencies of morphologically simplex words did facilitate lexical access for all frequency ranges. However, there are several issues with this analysis. First, a threshold for inflected words does not inform us about lexical listings for derived words and the morphemes they contain. Furthermore, Baayen et al. (1997) found base and surface frequency effects in Dutch well below the proposed 6 parts per million word threshold. This study also looked at monomorphemic words. They investigated the roles of word frequency itself, as well as the frequency of morphologically related words. They found that nouns with high-frequency plural forms have faster RT in a lexical decision task than nouns with lower-frequency plural forms. The frequencies of the words in the family did not have an effect, but the morphological family size did. Results of timing suggest the family size effect arises later in identification, not during earlier processing.

In support of a hybrid model, in which there is ample evidence that not only do words have a salient morphological reality, but also that morphological units play a role in the lexical access process, Meunier and Segui (1999) performed a study on auditory lexical access. They investigated the role of surface and cumulative (base) frequency in multimorphemic words. Multimorphemic words were found to be influenced by their surface frequency, and high and low surface frequency words were also affected by their base frequency. Furthermore, they suggested that suffixed words which belong to the same family are organized based on surface frequency and compete with one another during lexical access. When a multimorphemic word is accessed, the root is accessed, and also the affixes with which that root may be combined are accessed. These different combinations are what compete during the lexical access process. This results in inhibition (significantly slower RTs in a lexical decision task) so candidates with a well-populated morphological cohort have slower recognition than candidates with a sparsely-populated morphological family.

Additionally, López-Villaseñor (2012) looked at the effect of base frequency in multimorphemic words with derivational affixes. He found that base frequency did not affect RTs to visually presented Spanish words when affix productivity was low. This study showed that multimorphemic words with unproductive affixes are not processed in the same way as words with productive affixes, as words with productive affixes were decomposed and base frequency did have an effect. This study followed earlier work by Bertram et al. (2000a) that investigated the effect of base frequency in complex Dutch words with affixes that varied in productivity. The base frequency effect did not occur in words with low productivity affixes.

In returning to the Cohort model and associated models which incorporate a uniqueness point, Baayen et al. (2007) noted that the uniqueness point of a suffixed word is its non-suffixed counterpart. Baayen et al. (2007) investigated low-frequency multimorphemic words in two separate kinds of tasks across two modalities: naming and visual lexical decision in both the visual and auditory domain. They found that surface frequency was a significant predictor of RT, but base frequency was not. They also found an important relationship between a target word and its competitors. Morphological competitors were facilitatory and phonological relatives were inhibitory. Morphological competitors in their study included related words which shared a morphological connection, such as ‘hat’ and ‘hatter’. In contrast, those which had a phonological relation but did not share morphological relationship were words such as ‘hat’ and ‘chat’. They also found that low-frequency words are facilitated by greater affix productivity. The model proposes that words and morphemes are not predicted to be at different levels.

Finally, studies from neurolinguistics have found evidence of hybrid models of lexical access. Evidence for dual-route recognition is found in Lewis et al. (2011) as they presented evidence from Magnetoencephalography (MEG) data showing that participants who responded to words with a pseudo-affix showed effects of having both the whole form and pseudo-morphemes available for processing. Additionally, evidence from event-related fMRI in a study by Vannest et al. (2011) showed that some words are decomposable into their constiuent morphemes, such as ‘agreement’, and that others, such as ‘serenity’, are recognized whole-form.

1.1.4. Lexical access summary

In summary, three basic model types have been proposed that entail various incorporations of morphological structure in the lexical access process. Whole-word models represent the holistic storage of words and posit that a complex word is recognized holistically without decomposition into its constituent morphemes. In contrast, obligatory decomposition models suggest that complex words are decomposed into their constituent morphemes during lexical access, as evidenced by morpheme effects which operate independently from the whole-word. Finally, hybrid models allow for multiple pathways of recognition in which complex words are either recognized holistically or by being decomposed into their constituent morphemes. As outlined above, it has been suggested that affix productivity determines whether a word is recognized holistically or by its morphemes. In the current study, I explore the role of affix productivity in complex Arabic verbs. Specifically, I investigate the possibility of a hybrid model, and whether less productive affixes trigger holistic storage and more productive morphemes trigger decomposition. First, I will provide an overview of Arabic morphology below in section 1.2.

1.2. Arabic verb structure

The Arabic language comprises many varieties. One dialect, Modern Standard Arabic (henceforth MSA) is largely used for education, law, media, religion, and other official purposes. Another colloquial dialect is used for day-to-day usage. Colloquial dialects differ from both MSA and one another in non-trivial ways across their phonology, morphology, syntax, and lexicon, and these differences can result in a loss of intelligibility across dialects. Although colloquial dialects of Arabic are categorized into broad geographic terms based on general regions of usage, dialects can differ across many sociolinguistic dimensions, including the existence of urban-rural dialect distinctions (Holes, 1996; Miller, 2007), as well as dialect markers across gender lines (Abu-Haidar, 1989; Daher, 1998).

Given the fact that colloquial and MSA are largely restricted to specific domains, Arabic is typically described as being in a state of diglossia, in which the primary dialects of usage are utilized in addition to a highly standardized superposed dialect typically learned via formal education (Ferguson, 1959). An older codified variety, Classical Arabic contributes features to modern usage despite the fact that it, along with MSA, is not the native dialect of any one speaker or group of speakers (Ferguson, 1959). In these terms then, MSA can be considered the (H) dialect and the colloquial dialect would be the (L) form (Holes, 2004). The H dialect has traditionally been referred to as written Arabic and the local L dialect as spoken Arabic because MSA has been the form codified by standard written usage in schools and media. However, in recent years this distinction has been insufficient as colloquial dialects have begun to be written down in great quantities due in part to the spread of social media and informal text-based communication via internet or telephone. The script used to write colloquial Arabic differs based on region, mode of communication, and even from person to person (Warschauer, 2002; Palfreyman and Khalil, 2003) but typically speakers write colloquial text using much the same script as is used for MSA, with some modifications (see chapter 2 for detail).

Frequent mixing of the (H) and (L) varieties has led many to posit the existence of a dialect that can be termed Educational Spoken Arabic, which can be described as colloquial Arabic with heavy MSA elements that can be widely understood across speech communities (Mitchell, 1986, 1978; Bassiouney, 2009). Furthermore, it has been proposed that considering each dialect—be it MSA or colloquial or some combination—discrete entities that speakers code-switch between fails to characterize actual language utterances of Arabic speakers and that a continuum with Classical Arabic on one end of the spectrum and uneducated colloquial on the other on is a more appropriate illustration of Arabic usage (Badawi, 1973; Ryding, 1991; Kaye, 1994; Hary, 1996).

When categorized in a coarse-grained manner by region, colloquial Arabic dialects can be categorized into five major regional varieties: North African (spoken in Morocco, Libya, Tunisia, Algeria, and Mauritania), Egyptian, Iraqi, Gulfi (spoken in Saudi Arabia, Qatar, the United Arab Emirates, Bahrain, Oman, Kuwait, Yemen), and Levantine (spoken in Syria, Lebanon, Palestine, and Jordan).

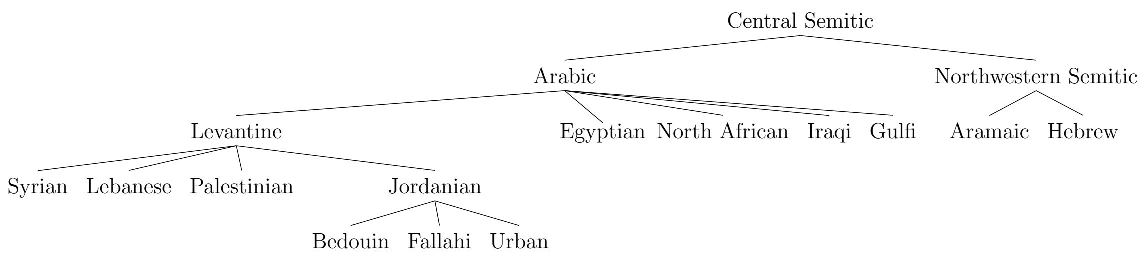

Figure 1.1: Linguistic family tree positioning Jordanian[1]
(expanded from Hetzron and Bender (1976))

For the work undertaken in this dissertation, Jordanian was selected as the dialect to be studied. Jordanian is a subdialect of Levantine, and further comprises three subdialect groups: Bedouin (rural), Fallahi (rural), and Urban (Suleiman, 1985). This is depicted in relation to other Semitic branches in Figure 1.1. Historically, the dominant dialect group in Jordan is Urban (Al-Wer, 2007). In the twentieth century, Jordan saw increasing levels of urbanization as populations moved from rural areas into the cities, and as other populations moved into the Jordanian capital Amman, including 1.5 million displaced Palestinians between the years 1948 and 1967 (Al-Wer, 2003, 2007). By 2003, one-third of Jordan's population lived in Amman, further cementing Urban as a dominating dialect of the country (Al-Wer, 2007). The dialect of focus in this dissertation is the current dominant Urban dialect spoken throughout Jordan and centered in Amman. For simplicity's sake, and to be consistent with previous studies on this dialect, it is referred to as simply "Jordanian" throughout the remainder of the dissertation.

Jordanian verbal structure is typical of other Semitic languages. Semitic morphology is characterized by its utilization of both concatenative and nonconcatenative processes. Many theories of the Semitic morphological system propose that the main

[1]This diagram is intended to situate the target dialect in the reader's mind, and not as a complete illustration of the relations between dialects. For example, it can be claimed that the Yemeni dialects are outside of the Gulf division given their substantial differences when compared to other Gulfi dialects such as Saudi (Watson, 2007). However, issues such as this are beyond the scope of this dissertation.

building block of the word is composed of a root which is typically triconsonantal but may take other forms, and a pattern which usually contains vocal elements but may also contain consonants. Both the root and the pattern are non-contiguous bound morphemes, meaning they are unpronounceable units unable to stand underived on their own, and they encode different kinds of information. Whereas a root encodes broad semantic meaning, a pattern provides non-referential semantic information such as mood, aspect, tense and inflection, as well as derivational information such as part of speech (Cantineau, 1950; Cohen, 1951). The canonical example of the root and pattern system is the combination of root /ktb/ which denotes the broad semantic sense of 'writing' with verbal pattern /CaCaC/ (where *C* indicates the root consonants) to become [katab] 'he wrote' and with nominal place pattern /maCCaC/ to form [maktab] 'office'. (Contemporary accounts of root-pattern morphology in Semitic include (Hilaal, 1990; Prunet et al., 2000; Versteegh, 1997; Davis and Zawaydeh, 2001). See (Berent et al., 2007; Ratcliffe, 2004; Ussishkin, 1999, 2005) for dissenting stem-based theories, although a fine-grained distinction between "root" and "stem" is not within the scope of the current study, and has no bearing on the claims made within.)

MSA exhibits 10 verbal paradigms[2] with each paradigm utilizing a pattern. Despite the English usage of 'measures' and the existence of the Arabic term 'awzaan' to refer to these verbal paradigms, for the purposes of consistency with previous studies on Semitic languages, in this dissertation I adopt the Hebrew term 'binyan' (plural: 'binyanim') to refer to these verbal paradigms, following McCarthy (1981). Binyanim are conventionally referred to by Roman numerals, for example Binyan VII denotes verbs of the form /nCaCaC/. Each binyan typically provides specific semantic meaning, for example Binyan II verbs tend to be either the causative or intensitive version of a Binyan I verb of the same root consonants. For example, the root /drs/ is asso-

[2]Classical Arabic contained as many as 14 binyanim, but a minuscule number of verbs from these other four binyanim have been attested in modern dialects. For instance, the usage of Binyan XII verbs of the form /iCCawCaCa/ to indicate colors and defects (Jastrow, 1980).

ciated with semantic sense of learning, and a derivation of this root in combination with Binyan I /CaCaC/ results in [daras], 'he learned'. In comparison, a derivation of this root with the Binyan II form /CaC:aC/ results in the word [dar:as] meaning 'he taught'.

The attested binyanim of Jordanian are shown in Table 1.1, with the typical vocalic patterns for each. Words from each binyan are shown in the right-most column for illustrative purposes.

Binyan	Form	Example
I	CaCaC	ʔaʕad "sat"
II	CaC:aC	sak:ar "closed"
III	CaaCaC	naaʔaʃ "discussed"
IV	ʔiCCaC	ʔitʕab "make tired"
V	itCaC:aC	itdʒan:ab "avoided"
VI	itCaaCaC	itdʒaahal "ignored"
VII	nCaCaC	nbasatˤ "was happy"
VIII	iCtaCaC	iftataħ "opened"
IX	iCCaC	iħmar: "reddened"
X	staCCaC	staxdam "used"

Table 1.1: Jordanian Binyanim (Verbal Paradigms)

Although all 10 binyanim are attested in MSA, distributions of binyan usage and productivity differ across dialect. For example, while Binyan VIII is very common in Sana'ani, a dialect spoken largely in the capital of Yemen, it is uncommon in Cairene, the dialect utilized in the capital of Egypt (Watson, 2007). Furthermore, Binyan VII may have limited productivity in Palestinian (Laks, 2009)[3]. However, some distributions are consistent across dialects. For example, Binyan IV is rare in all modern dialects of Arabic (Smolka et al., 2009), and has fallen out of productive usage in Maltese entirely (Kaye, 2007).

[3]A potential problem with the analysis Laks (2009) provides concerns the usage of the Elihay (2004) dictionary as a source for productivity across binyanim; however, native speaker judgment on contemporary usage contradicts this analysis (Farwaneh, personal communication).

1.3. Semitic studies in word processing

As outlined in the previous section, a considerable body of research has studied the contribution of frequency to morphological decomposition during spoken word recognition. However, frequency has been studied mostly in English and other Indo-European languages while research on Semitic languages has largely focused on other contributions to access, such as semantic and morphological features (Boudelaa and Marslen-Wilson, 2010; Frost et al., 1997, 2000; del Prado Martín et al., 2005). Furthermore, the vast majority of studies (with exceptions: Boudelaa and Marslen-Wilson (2013); Schluter (2013)) have been conducted in MSA, a domain-restricted dialect for which it can be argued there are no true native speakers (Ibrahim, 1983; Ferguson, 1959).

However, there has been extensive research regarding lexical access for Semitic languages focusing on the visual domain. Studies on Hebrew have demonstrated the psycholinguistic saliency of the root as the driving influence of word recognition. In a visual masked priming task, it was shown that prime-target pairs which shared the root resulted in facilitated lexical access as evidenced by faster RTs for lexical decision on the target word (Deutsch et al., 1998). Similarly, in Maltese, facilitation was demonstrated in a visual masked priming experiment for prime-target pairs which shared roots (Twist, 2006). Of most relevance to the current study is the research concerning Arabic. For visual word processing in Arabic, it has been shown that prime-target pairs which share roots in visual masked priming experiments result in a facilitated response to the target (Boudelaa and Marslen-Wilson, 2010, 2011, 2005, 2000, 2004b,a). In addition, in the neurolinguistic realm, magnetoencephalography (MEG) research has demonstrated that increased activity in superior temporal regions is consistent with models utilizing the root as an identifier, suggesting that roots are the units by which spoken words are recognized (Gwilliams and Marantz, 2015).

However, the visual domain is a troublesome modality to explore processing and lexical storage of Semitic words. This is due to the fact that most Semitic writing systems are unbalanced in the usage of consonants and vowels. In both Hebrew and Arabic, only consonants and a small subset of vowels are obligatorily written. Written passages which use encode all consonants and vowels are restricted to specific uses such as books for religious purposes, or those intended for beginning readers. For example, the written string كَتَب [katab] means ‘he wrote’ and كُتُب [kutub] means ‘books’, however the diacritics َ and ُ are the symbols which encode the vocalic elements and are not typically written. Thus, both words in typically usage would have the identical graphemic form كتب. This means that the orthographic form of the word is biased to representing the root and little to no pattern information is immediately discernible.

Furthermore, although colloquial Arabic dialects are typically written using MSA orthography (see chapter 2 for more details), the unfamiliarity associated with reading colloquial words in isolation could obscure target effects in a lexical decision task by forcing the participant to engage in too artificial a task. In the visual form it is not often clear if a written word in isolation is intended to be read as MSA or colloquial, which is an ambiguity that is immediately resolved when a word is pronounced in a colloquial pronunciation.

Therefore, the auditory domain is the most appropriate for an investigation into the effects of morphological structure on lexical access for speakers of Arabic. In concert with visual priming experiments, auditory priming experiments in Arabic have demonstrated a consistent effect of facilitatory root priming, and much of this research has been done in colloquial dialects. For Tunisian Arabic, Boudelaa and Marslen-Wilson (2013) showed prime-target pairs which shared roots resulted in facilitated RT of the root word. For Moroccan Arabic, Schluter (2013) showed that prime-target pairs which shared roots also resulted in faster RTs, using an auditory masked priming technique developed by Kouider and Dupoux (2005) intended as am analogue for visual masked priming and entailing durationally reduced primes embedded in auditory 'masks' composed of speech that is reversed and compressed. Other studies of Semitic auditory word processing are consistent with the Arabic results. Using an auditory masked priming experiment, Ussishkin et al. (2015) showed that root-sharing prime-target pairs in Maltese also successfully prime. Together, these results demonstrate a consistent saliency of the Semitic word root. By this, I mean that the results suggest the root as a unit of organization in the lexicon available to a speaker during lexical access.

Morpheme productivity presents an additional dimension for lexical processing and word recognition. Within Arabic psycholinguistics, Boudelaa and Marslen-Wilson (2011) investigated the roles of both root and binyan productivity in both visual masked priming and cross-modal priming tasks and found that root productivity was the primary driver of word decomposability during the lexical access process, as evidenced by facilitated RT when prime-target pairs shared a root. This was true even when the binyanim were unproductive. However, the reverse was not found to be true: even very productive binyanim did not result in facilitated lexical access when prime-target pairs shared binyanim but the target contained an unproductive root. They define morpheme productivity in terms of family size, which is the number of unique words which contain that morpheme.

1.4. Focus of the dissertation

Although there has been extensive research within the domains of both spoken and visual word recognition which argues for the complete decomposability of the Arabic word and the salience of the root (Boudelaa and Marslen-Wilson, 2004b, 2005, 2011), no study currently draws on the evidence for interactions between surface and base frequency to determine if there are different levels of decomposability. Given that it has been demonstrated Arabic words are readily decomposed into their roots during lexical access, what role does the binyan play? Does the productivity of the binyan determine word decomposability, as has been demonstrated for English, Dutch and Spanish? As outlined above, López-Villaseñor (2012), among others, demonstrated that in a lexical decision task, morphologically complex words with more frequent bases are responded to more quickly only when the derivational affix was productive. For words with unproductive derivational affixes, the frequency of the base morpheme did not have an effect on recognition speed. In contrast, these words were processed at a speed depending on the frequency of the surface word as a whole. This suggests that word recognition operates via a hybrid model that allows multiple strategies of storage and access of multimorphemic words. One path accesses words holistically without decomposition of the word into its constituent morphemes, and another path decomposes words into their constituent morphemes. Whether or not a morphologically complex word containing a derivational affix is recognized holistically or via decomposition depends on the productivity of that affix. Therefore, the questions at the heart of this study are: Does Arabic support a hybrid model in which some words are stored holistically and others are decomposed? Alternatively, are all words stored whole-form, or do all words undergo automatic decomposition into a root and pattern?

This dissertation aims to provide a more complete understanding of how complex word processing is affected by morpheme frequency, particularly for Arabic. Despite the extensive work showing morphological decomposability in Arabic, it remains unclear how morpheme frequency affects decomposition during lexical access. Specifically, does the frequency of the root morpheme facilitate speed of lexical access for all words, or does the productivity of the binyan determine whether or not the word is accessed holistically? Furthermore, the effects of morpheme frequency in a nonconcatenative system, particularly with discontiguous morphemes, are understudied. This study investigates the question of pattern productivity and its role in lexical access in Jordanian (specifically Urban Jordanian/Ammani), a colloquial Levantine dialect which has been under-studied in psycholinguistics, despite having been well-studied phonologically and morphologically (Davis, 2011). Jordanian utilizes multiple binyanim of varying productivity, and the current study investigates how binyanim which differ in their relative productivity affect the lexical access process. Much of the previous investigation into morphological decomposability of Arabic words has utilized cross-modal priming and masked priming (Marslen-Wilson et al., 1994; Gaskell and Marslen-Wilson, 2002; Marslen-Wilson et al., 1996). Cross-modal priming has been shown to elucidate morphological decomposability such that only words with true morphological relationships prime related targets (Marslen-Wilson et al., 1994). However, cross-modal priming requires a visual component. Because of the existence of lexical overlap between colloquial dialects and Modern Standard Arabic, and because most features of colloquial pronunciation are not indicated in orthography, a speaker exposed to a visual stimulus of a colloquial-MSA cognate may activate their knowledge of MSA and not their colloquial dialect. Furthermore, MSA's limited-domain usage and the existence of age of acquisition effects in lexical processing (such as those in Nagy et al. (1989), among others) may obscure effects associated with morpheme frequency and productivity.

Auditory lexical decision provides a relatively replicable method for measuring on-line processing Goldinger (1996). Although auditory masked priming avoids the potential confound of MSA - colloquial confusion, masked priming illuminates issues of early processing and may then measure indiscriminate decomposition early in the lexical access process regardless of true morphological relationship, as in the potential decomposition of pseudoword 'snay' and the past morpheme '-ed' in 'snay/snade' (Tyler et al., 2002) which means it is unlikely to inform any possibilities of wholeform storage. Finally, because regular lexical decision without priming is sufficient to show morpheme frequency effects, it is unnecessary to use priming in this study. Auditory lexical decision tasks investigating the roles of base and surface frequency are an effective way to look at effects in Arabic without resorting to written form, and without over-generalizing the decomposability of the word into its constituent morphemes. Therefore, the chosen modality for the word processing experiments in this dissertation is auditory lexical decision.

The results of four experiments addressing the role of morphological structure in lexical access and the relationship between decomposability and productivity are presented in this dissertation. These experiments were conducted with native speakers of Jordanian, and were conducted in Amman, Jordan. Experiments 2a and 2b focus on a productive derivational affix corresponding with a verbal paradigm typically referred to in Semitic studies as Binyan I. Experiment 2a shows a significant facilitatory effect of the frequency of the root morpheme on the speed at which speakers recognized a Binyan I verb, and results from Experiment 2b provide evidence that surface frequency also has a facilitatory effect on the speed of recognition for these verbs. In contrast, Experiment 3a shows that words formed with an unproductive derivational affix corresponding to the paradigm Binyan VIII do not exhibit a facilitatory effect of base frequency of the root, and only the frequency of the whole word determines the speed at which speakers recognize that word. Experiment 3b provides further evidence that decomposability is tied to productivity, as another unproductive paradigm, Binyan X, is populated by verbs for which the base frequency is irrelevant during auditory lexical access as only the surface frequency of the word facilitated recognition speed.

Collectively, these results support a hybrid model of auditory word recognition in which morphologically complex words are recognized via multiple potential pathways which either decomposes the word into its constituent morphemes or recognizes the word holistically in its derived form. Whether or not a word is recognized by the morphemes it contains or by the whole-form depends on the productivity of the derivational affix, supporting conclusions proposed by Baayen et al. (2007); Bertram et al. (2000a,b); López-Villaseñor (2012); Taft and Nguyen-Hoan (2010) in which affix productivity determines the decomposability of a word during lexical access.

1.5. Organization of the dissertation

This remainder of this dissertation is organized as follows: chapter 2 describes the construction of a corpus of a colloquial Jordanian Arabic and Experiment 1, a word familiarity experiment; chapter 3 presents and discusses the results of Experiments 2a and 2b, auditory lexical decision experiments investigating the roles of base frequency of the root morpheme and surface frequency of the word for a verbal Arabic paradigm. Chapter 4 contains Experiments 3a and 3b, which also explore the roles of base and surface frequency, but for unproductive verbal paradigms of Arabic; and finally chapter 5 offers general discussion, discusses implications of the experimental results, and concludes the dissertation.

1.6. Conclusion

As outlined in this introductory chapter, this dissertation contributes to issues in Arabic psycholinguistics by addressing the question of how root frequency affects speed of recognition is productive vs. unproductive binyanim. Although numerous studies have demonstrated that Arabic words are decomposable during lexical access as evidenced partly by the psycho-saliency of the root in priming experiments, the role of binyan productivity has only begun to be explored. In this dissertation, I investigate the effects of both surface frequency of a whole word and the base frequency of the word's root morpheme in both productive and unproductive binyanim in Arabic, and provide evidence that a binyan's productivity determines whether or not the base frequency of the root morpheme is a relevant factor in determining speed of lexical access during auditory word recognition.

2. Corpus construction and Experiment 1

This chapter provides details of the creation of a corpus of Jordanian Arabic to serve as a resource for word frequency and morpheme frequency measurements as well as the productivity measures for the experiments presented in chapter 3 and chapter 4. Here, I present some of the methods utilized during construction of the corpus. The corpus construction project is part of an ongoing effort to construct a balanced, representative corpus of Jordanian Arabic. Brysbaert and New (2009) explored reliability of frequency measures from various corpus sizes, and determined that for high frequency words, a corpus size of 1-3 million words is sufficient, but for low frequency words with frequencies of lower than 10 parts per million, a corpus size of at least 16 million words is sufficient. This chapter details my efforts in creation of a 16-million word corpus sufficient for lexical statistic measurements formed from subsets of previously existing resources. In this chapter, I provide an overview of some of the challenges of creating and using corpora in colloquial Arabic dialects. Then, a description of the consolidation process I employed to aggregate existing Jordanian corpora and use machine learning to determine which portions of existing Levantine corpora are Jordanian dominant is offered, as well as details of efforts to augment these resources with the creation of a new Jordaninan corpus. After discussing the methods for creating the corpus, I present Experiment 1, a word familiarity experiment performed to test the validity of the completed corpus. I show through the resulting correlation between corpus frequency and word familiarity that the corpus is a representative sample of Jordanian. Finally, I provide an overview of corpus-based productivity measurements of the Jordanian binyanim (verbal paradigms) discussed in chapter 1.

2.1. Colloquial Arabic corpora: an overview

As discussed in chapter 1, Modern Standard Arabic (MSA) is the dialect of official media and education. MSA is the only dialect with a written standard, and the colloquial dialects thus lack standard orthography.

Traditionally, MSA has been referred to as written Arabic and colloquial dialects as spoken or vernacular Arabic (Ferguson, 1959). However, in recent years colloquial dialects have begun to be written down with great regularity in part due to the spread of social media and informal text-based communication via internet or mobile phone (Bakr et al., 2008), which means that colloquial dialects are producing enough text to make corpus-based research feasible. Despite this, resource creation for colloquial Arabic dialects has been limited in both size and scope (Diab and Habash, 2007; Habash et al., 2013; Pasha et al., 2014). Some of the challenges facing creation and utilization of colloquial Arabic corpora stem from its lack of standard orthography. These challenges exist in addition to the features already presented in MSA resources such as the high degree of homography resulting from the fact that Arabic short vowels are represented by optional diacritics in the Arabic writing system, meaning that the string يعد /jʕd/ could potentially be a graphemic representation of 'return', 'promise', 'count', 'repair', among other words.

In addition to pan-Arabic challenges of corpora utilization, corpora of colloquial Arabic have additional features complicating their creation and usage. (For a complete overview of some of the script-specific issues surrounding colloquial usage in text, see (Eskander et al., 2013).) Despite the fact that colloquial Arabic is being written down in abundance, the lack of writing standard results in a highly variable spelling system. For example, the future marker can be written using either the letter ح /ħ/ or ه /h/ in Egyptian colloquial writing, and these two symbols are in free variation with no apparent factor determining usage of one vs. the other (Eskander et al., 2013). Another common feature of colloquial writing is the omission of standard-usage double dots on some letters, collapsing them with similar graphemes. Thus, although MSA has two separate letters for ه /h/ and ة /a/, typical colloquial texts write both letters as ه /h/ which can also be observed in the colloquial usage of ى /a/ to represent both letters ى /a/ and ي /j/ as they are typically written in MSA (Buckwalter, 2007). Finally, the MSA letter ء which is one of the script encodings of /ʔ/ is frequently omitted in colloquial text (Diab and Habash, 2014), likely due to the frequent reduction of /ʔ/ in spoken colloquial Arabic (Ferguson, 1959).

In addition to these regularly-attested alternations, the lexical differences between MSA and the colloquial dialects result in variable spellings (Diab and Habash, 2014). Consider Example 1, and note that for one word a., there is a variety of valid orthographic forms depending on if the writer chooses to split the initial part of the circumfix from the word, as in b., or attach it such as in d. and e., as well as if they chose to use letters ا /a:/ and و /u:/ to encode long vowels or analyze the sounds as short vowels and they are not reflected in the orthography. Crucially, none of the possible spellings in b.-e. are considered 'wrong' and would be accepted by writers of colloquial dialects. The nonstandard forms emerge from the fact that neither the stem or the negation circumfix exist in MSA, and thus writers are left to determine a spelling for themselves.

(1) a. *ma:-ʃuft-ʃ*
NEG-saw.1s-NEG
'I didn't see.'

b. ما شفتش *ma: ʃuftʃ*

c. ما شوفتش *ma: ʃu:ftʃ*

d. ماشوفتش *ma:ʃuftʃ*

e. مشوفتش *maʃu:ftʃ*

To cope with issues of orthographic inconsistency, normalization solutions have been proposed, chief among them the Conventional Orthography for Dialectal Arabic (CODA) proposed by Habash et al. (2012a), and developed for automatic normalization by Zribi et al. (2014), though other normalization systems have also been proposed, such as one focused on normalization of language used in microblogs (Tweets) by Darwish et al. (2012). These systems vary on what degree of variability is acceptable in the output, but they converge on similar goals: the conversion of raw non-standard, highly variable text into something more computationally robust.

However, even when orthography is normalized, several challenges remain in creating and utilizing colloquial corpora, including the high degree of lexical overlap between colloquial dialects as well as between the colloquial dialects and MSA. This can make dialect identification in text a non-trivial task. Mubarak and Darwish (2014) showed that lexical items which are highly colloquial and are not present in MSA may be used in multiple colloquial dialects and are therefore not good predictors of dialectal variety by region. For example, the word /e:ʃ/ 'what' is used in multiple regional varieties and although it may be a useful marker for classifying colloquial text vs. MSA text, it is not useful for dialect classification within the colloquial dialects.

Accompanying lexical cognates is the issue of lexical divergence. The divergent meanings between MSA and colloquial cognates, and between cognates from one dialect to another means that even in robust lexicons there is the potential to misguide machine translation systems or other components reliant on meaning. The same graphemic word can have highly divergent meanings across colloquial dialects, for example the word بلش /bal:iʃ/[1] means 'begin' in Levantine dialects, but means 'end' in Sudanese dialects (Diab and Habash, 2014). Similarly, العافية /alʕa:fija/ is commonly used to mean 'health' in Egypt and the Levant, but means 'hellfire' in Morocco (Diab

[1]The IPA transcription provided for the cross-dialectal examples here may differ in vowel quality between the dialects, as the purpose of these examples is to show homography and not phonemic similarity.

and Habash, 2014). Finally, ماشي /ma:ʃi:/ is a common affirmative utterance in Egypt and the Levant meaning ‘ok’ or ‘yes”, but in Yemen it is ‘no’.

Given the issues outlined here, it follows then that colloquial Arabic corpora have been relatively limited in scope when compared to MSA corpora.

For both MSA and colloquial Arabic, Arabic corpora pre-dating the web are limited. A rare example is the Buckwalter corpus (Buckwalter, 2003), which began with the digitization of newspapers in about 1986 and consisted of approximately 40,000 words; the Buckwalter corpus grew to 3 billion words with the emergence of the web, and is primarily used for lexicography, but is not available in its raw form.

Another corpus of MSA that is composed primarily of newswire text is the Arabic Gigaword (Parker et al., 2011), which is composed of nine newspapers across the Arabic-speaking world and contains approximately 400 million words.

As discussed in chapter 1, because Arabic as a language is characterized by usage along a spectrum of MSA and colloquial, frequent code-switching between a given speaker’s colloquial dialect and MSA results in the fact that corpora intended to be MSA-focused contain a measure of colloquial language. For example, Arabic Gigaword contains over 2,000 instances of the non-MSA Egyptian word كده /kida/ ‘like that’ as shown by Mubarak and Darwish (2014)

Similarly, colloquial corpora can exhibit MSA because of this code-switching. For example, Wray and Ali (2015) created an 850 hour multi-dialectal speech corpus from interview and debate programs from the Arabic broadcast channel Al Jazeera. These programs were chosen for their heavy colloquial content, but Wray and Ali (2015) found that speakers produced MSA-dominant utterances 28% to 53% of the time even as bi-dialectal code-switching resulted in their very next utterance being in a colloquial dialect.

In addition to the Al Jazeera corpus, additional colloquial corpora constructed using broadcast recordings include the Qatari Arabic Corpus (Elmahdy et al., 2014), a 15-hour corpus created by manually transcribing colloquial talk shows and series. Other wide-band/high definition speech corpora include the A-SpeechDB corpus (ELRA, 2014), which contains approximately 20 hours of Egyptian intended to provide coverage of phonetic phenomena by a multitude of speakers. Another specialized audio corpus is the BBN-Darpa Babylon corpus which contains 45 hours of Lebanese focused on medical training speech (Makhoul et al., 2005). Prompted audio selections in Levantine, Egyptian, and Gulf dialects make up the 32-hour Multi Dialect Arabic Speech Parallel corpus (Almeman et al., 2013), and finally the OrienTel-Telephony Database (Iskra et al., 2004) is a multi-dialectal corpus that consists largely of directory information, numbers, dates and times, and other non-spontaneous speech suited for specialized usage.

Pre-dating the creation of these audio corpora, narrow-band speech corpora at 0-4k Hz were created largely via telephone. As explored further below in section 2.2, two telephone corpora exist for Levantine: Fisher (Maamouri et al., 2007) and Appen (Appen, 2007). Both corpora were created by having native speaker contributors call an acquaintance and discuss a pre-assigned topic. The resulting spontaneous conversation was then professionally transcribed. These corpora creation methods were also utilized for other colloquial dialects, including for the Egyptian Callfriend (Canavan and Zipperlen, 1996) and Egyptian CallHome (Canavan et al., 1997) corpora.

In addition to colloquial corpora constructed from audio and the the accompanying transcriptions, efforts in colloquial text corpus building have begun to become more robust in the advent of increased colloquial usage on the web. McNeil and Faiza (2011) created a corpus of Tunisian colloquial Arabic by not only incorporating traditional colloquial texts such as television scripts and folktales, but also by mining web platforms which exhibit frequent production of colloquial text such as Facebook and blogs. Another example of web mining to create a corpus of dialectal Arabic is the Arabic Online Commentary (AOC) dataset (Zaidan and Callison-Burch, 2011b; Cotterell and Callison-Burch, 2014). Although newspaper articles themselves are written in MSA, the comments provided by readers in the commentary sections of online newspapers are largely colloquial. Zaidan and Callison-Burch (2011b); Cotterell and Callison-Burch (2014) scraped comments from five major newspapers located in different regions to obtain a written material from MSA in addition to Iraqi, Levantine, Egyptian, and Gulf dialects, resulting in a 52 million word multi-dialectal corpus, of which approximately 41% contains colloquial material. In addition to newspaper commentary, colloquial text is heavily used in 'microblogs' such as those written using the Twitter platform. YADAC, an Egyptian corpus of 6 million words, was created by harvesting microblogs as well as forums, blogs, and online knowledge platform sites where users provide both questions and answers and are typically informal in nature (Al-Sabbagh and Girju, 2012). Microblog retrieval was also utilized by Mubarak and Darwish (2014), who created a multi-dialectal Twitter corpus by exploiting user-provided information about location as well as geo-coded information which indicates absolute position. Finally, there is the strategy of using words typically used in one dialect and utilizing a search engine to explore potential sources of colloquial text. Almeman and Lee (2013) performed a survey of speakers from Gulf, Levantine, Egyptian, and North African dialects to determine lists of colloquial words to bootstrap collection of web corpora.

In sum, effective methods for corpus creation of colloquial material have been demonstrated in the forms of web mining for written colloquial text, transcription of broadcast audio, and inclusion of traditional materials. Although there are some existing Levantine colloquial corpora, they are largely mixed material from not only Jordanian but also Palestinian, Syrian, and Lebanese. In section 2.2, I describe the process of consolidation of existing Jordanian corpora and and of developing strategies of classifying mixed Levantine corpora into subdialects in order to isolate Jordanian material. Following this, I present a new corpus in section 2.3 and describe the methodology utilized in corpus creation.

2.2. Consolidation of existing resources

As discussed above, the goal of creating a corpus for this dissertation is to have a resource to provide lexical statistics necessary to perform a psycholinguistic study focusing on word frequency, morpheme frequency, and affix productivity. Following Brysbaert and New (2009), I aimed for approximately 16 million words as a reasonable corpus size. However, previously existing corpora (Almeman and Lee, 2013; Maamouri et al., 2007; Zaidan and Callison-Burch, 2011b; Mubarak and Darwish, 2014; Appen, 2007) total only approximately 11 million words. Furthermore, only a small subset of these corpora (specifically, Maamouri et al. (2007); Mubarak and Darwish (2014); Appen (2007)) contain fine-grained enough annotation to determine what material is specifically Jordanian: the remaining corpora include Jordanian together in a collection with other Levantine material.

The Fisher Levantine corpus (Maamouri et al., 2007) is composed of transcripts of 279 telephone conversations (total 45 hours, which when transcribed equals approximately 650,000 words.) The corpus was collected via a collection platform which contacted participants and assigned them a particular topic to discuss. Topics rotated every 24 hours. Each call transcript is annotated with information for each speaker's native dialect, as well as age and gender. The Appen Levantine Corpus (Appen, 2007) was constructed in an identical manner, and offers an equal level of speaker annotation. Using these speaker annotations, I was able to have sentence-level classifications for each utterance in the corpus, and have datasets composed of each Syrian utterance, each Lebanese utterance, and so forth. A total breakdown of the combined speech corpora is shown in Table 2.1.

In addition, some speech collections (such as the Jordanian colloquial portion from Iskra et al. (2004)) were not deemed suitable for inclusion in the corpus as they contained largely directory information, numbers, dates and times, and other non-spontaneous speech suited for specialized usage. Thus, their inclusion may artificially

Dialect	Transcript word count
Jordanian	250,915
Palestinian	142,207
Syrian	132,417
Lebanese	121,193

Table 2.1: Content of Fisher (Maamouri et al., 2007) and Appen (Appen, 2007) corpora by dialect

alter final lexicality statistics.

To target Jordanian as a subdialect specifically and exclude other Levantine dialectal material, I performed dialect classification on the aggregated Levantine corpora. The goal of dialect classification is to automatically identify the dialect of a sentence after training a classifier on manually-labeled material for each dialect. I trained a series of language models, probability distributions of sequences of words and characters, to provide automatic sentence-level classifications for dialect of one of four sub-Levantine varieties: Jordanian, Syrian, Palestinian and Lebanese.

In the text domain, previous work on dialect identification has been performed for purposes ranging from training automatic speech recognition to bootstrapping automatic corpus building efforts. For some of these studies, the focus is on a binary classification between MSA and a colloquial dialect, such as the Egyptian-MSA classification performed by Tillmann et al. (2014); Mansour et al. (2014); Elfardy and Diab (2013). A more fine-grained classification is performed by Zaidan and Callison-Burch (2011b), as they crowdsource the human classification of 100,000 sentences of dialectal Arabic, and train language models to classify additional dialectal material scraped from online newspaper comment sections. The classification categories for this study were: MSA, Levantine, Egyptian, and Gulf. On this 4-way classification, Zaidan and Callison-Burch (2011b) achieve an accuracy of 69.4%. However, their accuracy when classifying between just the non-standard dialects Levantine, Egyptian, and Gulf, jumps to 83.5%.

Typically, dialect classification is performed into these coarse-grained categories. However, work by Malmasi et al. (2015) attempted classification with a finer-grained distinction. They classify text into seven categories: MSA, Tunisian, Egyptian, Jordanian, Syrian, and Palestinian, achieving a 74.3% accuracy for this 6-way classification.

However, a potential pitfall of this work concerns the materials used as the training dataset utilized in constructing the language model. They use a parallel corpus composed of the above 5 non-standard dialects in addition to MSA from (Bouamor et al., 2014), but the corpus was created under prompts in such a way that it is not entirely parallel. For example, when speakers who contributed to the corpus were prompted to translate a base phrase into their native dialects, some speakers inserted additional material (such as /maʃa al:a/ 'wow!') while others did not, despite these phrases being pan-Arabic and having no preponderance to occur more frequently in one dialect over another. It's likely then that these phrases emerge in the corpus as an artifact of a particular contributor's inclination to use them at that time.

In an attempt to utilize a spontaneously created training set, I harvested geo-tagged tweets and location-specificed from the countries most strongly associated with that dialect, and also included tweets collected by Mubarak and Darwish (2014). The purpose of using Twitter data is to help ensure that the language produced is entirely spontaneous as opposed to prompted.

2.2.1. Materials

I utilized 3,000 tweets per sub-dialect (Jordanian, Palestinian, Syrian, Lebanese) which contained approximately 30,000 words. I drew partially from annotated material from Mubarak and Darwish (2014) which contains Twitter data from Jordan, Palestine, Syria, and Lebanon and was collected by exploiting user-provided location and geo-tagging info, and I supplemented additional tweets harvested in the same manner from the 2014-2015 period. (See section 2.3.1 for a detailed account of tweet harvesting by geographic location.) Data was prepared for study by removing information included in the tweets such as emojis, other usernames (mentions), punctuation, and links.

2.2.2. Procedure

To determine what strings in each dialectal text most reliably distinguished it from the other dialects, the text was split into testing and training sets and language models developed on the training sets were evaluated on the testing sets, following standard practice. I used the SRI Language Modeling (SRILM) toolkit (Stolcke et al., 2002) for building and applying a series of language models encompassing the following:

- *Word n-grams:* Modeled the probability of occurrences of contiguous word strings in a text. For the current study, three word n-gram models were built: a unigram model determining the likelihood of a word occurring in a text, a bigram model determining the likelihood of a word occurring given the previous word, and a trigram model determining the likelihood of a word occurring given the previous two words. Unknown words encountered by the model were mapped to <unk>, and sentence start and end markers were considered in the models.

- *Character n-grams:* Modeled the probability of occurrences of contiguous character strings in a text. Three character n-gram models were built: a unigram model determining the likelihood of a character appearing in a text, a bigram model determining the likelihood of a character occurring given a previous character, and a trigram model determining the likelihood of a character appearing given the two previous characters.

Thus, six model types were built for each dialect.

2.2.3. Results

Classification accuracy was measured by calculating the perplexity (a measurement of how well the probability model predicted the testing set) of each dialect's language models on the testing set, and selecting the lowest perplexity across the four dialect models as the classification for that sentence. Following previous studies, I used k-fold cross-validation with k = 10. Using this method, a random sampling of 9/10 of the data was used as the training data with the remaining 1/10 being used as testing data, and this process was repeated 10 times. The accuracy reported below is calculated as an average across all 10 folds. I used a random baseline to compare accuracy results to. As there are 4 different categories of dialect, the random baseline was $1/4 = 25\%$.

Accuracy for different language models is shown in Table 2.2.

Language model	10-fold cross-validation accuracy
Character unigram	41%
Character bigram	57%
Character trigram	65%
Word unigram	67%
Word bigram	**69**%
Word trigram	68%

Table 2.2: Classification accuracy for speech training and testing set. Best results are in bold.

As shown in Table 2.2, the highest performing model was the word bigram model which achieved 69% accuracy compared to a 25% random baseline. The least informative probability distribution for identification of transcribed speech data was the unigram character model at 41% accuracy.

After determining the most accurate classification method based on the language models and corpora explored here, I used the bigram word model to classify the mixed Levantine corpora from Almeman and Lee (2013); Zaidan and Callison-Burch (2011b) to classify each sentence in the corpus. In the end, 41% of the sentences

Corpus	Total Words	
Fisher (Maamouri et al., 2007)		169,870
Appen (Appen, 2007)		81,047
Twitter (Mubarak and Darwish, 2014)		712,367
Corpus	Total Words	Classified Jordanian
Multi-Dialectal (Almeman and Lee, 2013)	10,833,667	4,294,785
AOCD (Zaidan and Callison-Burch, 2011b)	223,622	190,407
		Total: 5,448,476

Table 2.3: Corpora subset creation

from Almeman and Lee (2013) and 71% of the sentences from Zaidan and Callison-Burch (2011b) was classified as Jordanian. The high percentage of Jordanian material found in the dataset from Zaidan and Callison-Burch (2011b) is not surprising as the newspaper selected to represent Levantine for that multidialectal dataset was indeed a Jordanian newspaper. The final counts of the aggregated corpora are shown in Table 2.3.

2.3. Building a new corpus

To augment the 5 million word corpus resulting from the Jordanian subset of the mixed Levantine data described above, a new 11 million corpus was constructed from a variety of sources. First, I harvested web data written in Jordanian Arabic from Twitter, web blogs, and forums. Then, I selected videos from YouTube which exhibited heavy usage of spoken Jordanian Arabic and crowdsourced their transcription from native speakers. Finally, I digitized collections of published short stories and novels written by Jordanian and Jordanian-Palestinian authors.

2.3.1. Web-based corpus

The web-based portion of the new corpus was constructed in two ways. First, I harvested Twitter microblogs from Jordan. Then, I harvested web blogs and forums likely to contain Jordanian content.

2.3.1.1. *Twitter harvesting*

There are approximately 15 million tweets produced in Arabic per day (Mubarak and Darwish, 2014). Each tweet produced has associated metadata that includes the time of publication, the username of the author, the language the tweet is written in, as well as location information of where the tweet was written from. This metadata is searchable and made available for harvesting via the Twitter API[2]. Given that all colloquial dialects of Arabic utilize the same script for encoding language, and because the writing of colloquial language is not a phenomenon unique to Jordan, the approach used in previous studies such as (Ljubešic et al., 2014) of harvesting all tweets the API returns as having been written in a specific language will not accurately target Jordanian text. A selection of 'Arabic' would produce all colloquial dialects, as well as material written in MSA.

As for location information, it comes in two forms. The first is a raw latitude longitude geo-tag which can give precise indication to region. However, this is an optional feature most users do not use: only about 2% of tweets are geo-tagged for location (Huck et al., 2012). Another optional feature is a user-provided location, which approximately 70% of tweets include (Mubarak and Darwish, 2014). These are locations which the user has custom written, and may or may not be informative as to actual location. For example, a user can write they are located in 'Amman, Jordan' or simply write 'My house'. To create the Twitter corpus, I used the Twitter API and engaged with it via the Twython library[3] for Python. Using these tools, I requested the Twitter API search through all tweets in the Arabic language which met either of the following criteria:

- Geo-tag: tweet was tagged within the geographic borders of Jordan

[2] https://dev.twitter.com/
[3] https://github.com/ryanmcgrath/twython

- User location: tweet was written by user with location matching Amman, as well as the country name itself. Following Mubarak and Darwish (2014), possible user location names associated with Jordan were drawn from the geographic name database GeoNames[4]. In addition, each time a user from Jordan was identified, manual inspection of user locations of their followers was performed to collect new permutations of possible ways to indicate residence in Jordan.

This approach operates under a couple of assumptions. First, that users tweet from a location where speakers speak the same dialect as the user. Given patterns of immigration in the Levant, as discussed in chapter 1, users who produce geo-coded tweets in Jordan may have origins, and therefore native dialects, associated with a different region. Furthermore, a user may be temporarily producing tweets with a different geo location than their default location if they are traveling. To partially circumvent this potential confound, tweets which exhibited mismatch between the user-provided location and the geocode were not retained in the dataset.

Tweets were collected for the greater part of a year to help reduce the possibility of cyclic effects of frequency, such as the tendency for holiday greetings to spike during Ramadan, for example (Eisenstein, 2013; Refaee and Rieser, 2014). To help maximize the likelihood that the collected tweets contained colloquial material, I discarded tweets which contained vowel diacritics, as this was typically indicative of a user tweeting a verse from the Quran. In the end, I collected 1,173,500 million words from Twitter. Material that was considered extra-textual, such as other usernames (mentions), emojis, or links, was discarded.

2.3.1.2. Web harvesting

In addition to scraping Twitter for text associated with the region of Jordan, I utilized the web as a resource for corpus building by formulating targeted queries of Jorda-

[4] geonames.org

nian content as well as exploited Google's explicit geographic marking of Jordanian websites.

Following previous work on colloquial Arabic corpus building (Almeman and Lee, 2013; Wray et al., 2014), text in the target dialect was identified and acquired by performing search engine queries of lexical items unique to that dialect. Wray et al. (2014) found that formulating queries with function words particular to that dialect yielded particularly robust results; however, since Jordanian shares function words with other Levantine dialects, these queries may result in mixed Levantine returns. Google Advanced Search offers an option to only return results from specific countries, however Wray et al. (2014) also found that websites which contained high sources of Yemeni content and drew large numbers of Yemeni readers were occasionally physically located outside of Yemen and were not found by this Google feature.

For these reasons, the identification of web sources of Jordanian text was performed in two ways: first, a list of common Levantine function words were queried while utilizing Google Advanced Search to target Jordanian websites. I selected 10 words, shown in Table 2.4. Second, a list of content words exclusive to Jordan were queried without the geographic targeting feature. Because most existing dictionaries such as (Elihay, 2004) focus on other Levantine dialects and are therefore not Jordanian-exclusive, the wordlist was pulled from the online dictionary Mo3jam[5]. This is a user-created lexicon focusing on colloquial Arabic dialects where contributors provide words along with definitions and regional dialect info. Other users may dispute or corroborate the information provided. I selected 10 words which were indicated to be in usage in Jordan, but were not attested in Syria, Lebanon, or Palestine. Given the informal nature of this resource, I selected words which were not currently in dispute with other users with regards to their meanings, existence, or exclusivity to the Jordanian dialect. Items which had sparked previous visible discussion on these matters were not selected. The words are shown in Table 2.4. Ten from each

[5] en.mo3jam.com

query were manually evaluated for Jordanian content. Queries were performed with dialect words in isolation, as well as paired with an another directing element (such as منتدى /muntada:/ 'forum' or the blogging platform 'WordPress'). Of the Jordanian words from Mo3jam which were queried without additional Google settings, 40% of the results evaluated were determined to have substantial Jordanian content. Of the Levantine function words paired with Jordan regionally-limited Google queries, 68% of the evaluated queries were determined to be Jordanian.

Jordanian words	Levantine words
zalimih 'guy'	miʃ 'not'
abi:ʃ 'there is not'	ʕam PROG
aðrib 'to like on Facebook'	raħ FUT
ʔanrad 'nerdiest'	ewih 'yes'
zaʕiʔ 'salty'	bid: 'to want'
daʃar 'to leave'	leʃ 'why'
$t^ʕ$aʔaʕ 'amazing'	ʃu: 'what'
zanʔaf 'to give/to leave'	eʃ 'what'
mudibris 'depressed'	ħada: 'someone'
laxmih 'stupid'	ʔiʃe 'thing'

Table 2.4: Words utilized for web queries seeking Jordanian content

In total, thirteen websites were determined to be robust sources of written Jordanian. I utilized the Beautiful Soup library in Python to scrape these sites and clean them of HTML and other extraneous formatting artifacts (Richardson, 2008). The breakdown of the web content, with the Twitter data described above, is shown in Table 2.5.

Source	Word count
Twitter	1,173,500
Forums (including Ana Jordan, Kooora)	2,877,000
Opinion columns (including Zad Jordan)	4,928,342
Blogs (including WordPress, Blogspot)	746,758
Total	9,725,600

Table 2.5: Scraped web content by source

2.3.2. Transcription of audio

Colloquial Arabic is found in abundance in audio and video media, from traditional broadcast to user-driven media platforms such as YouTube. However, considering the priority of the current study in morpheme and word frequency, raw audio is not of use to the current corpus without the accompanying transcription for automatic searchability and counting of words and word parts. However, professional transcription of colloquial Arabic is slow (4 days for 1 hour of speech) and expensive (US$300 for 1 hour of speech) (Wray et al., 2015). To overcome obstacles in resources, crowdsourcing is utilized as a viable alternative. Crowdsourcing is the process of breaking down a large task into smaller tasks and distributing the completion of these across many participants (the crowd). Crowdsourcing is not only a way to access large numbers of participants who are diverse both in demographics and skillsets, but also has become a standard for collection and annotation of data in numerous speech and language processing domains. For corpus creation, crowdsourcing has been shown to be an effective way of creating large corpora of transcribed speech in many languages, including English (Lee and Glass, 2011; Marge et al., 2010a,b; Hämäläinen et al., 2013), Spanish (Audhkhasi et al., 2011), Swahili, Amharic (Gelas et al., 2011), Korean, Hindi, and Tamil (Novotney and Callison-Burch, 2010).

Because contributors are non-professionals, quality control is a non-trivial component of crowdsourcing, and performing automatic online quality control rejects spam submissions and limits data attrition when compared to leaving quality control to post-processing. A typical form of crowdsource quality control is the utilization of the gold standard in which a percentage of the tasks is completed ahead of time by an expert. Contributor attempts are compared with the expert answers and either accepted or rejected accordingly.

Moreover, not every task is appropriate for the crowd. For example, a transcription task can be executed with non-professionals because it simply requires that a participant answer the question "What did you hear?" More complicated questions which require a specialized skillset or deep meta-linguistic knowledge is less appropriate for the crowd: for example, a study which required participants determine the case marking of an Arabic word was less successful (Zaghouani and Dukes, 2014). When proper quality control methods are executed, the quality of the resulting product is favorably comparable to professional quality (Zaidan and Callison-Burch, 2011a; Novotney and Callison-Burch, 2010). Furthermore, when compared to professional rates and speeds, the crowd is both faster and cheaper: 1 hour of colloquial Arabic speech when distributed in pieces to numerous non-professional native speakers is transcribed by multiple contributors in 18 hours and costs US$42 (Wray et al., 2015).

For the current study, colloquial-heavy programs available on YouTube were selected for transcription. Videos of multiple types were selected. First, the Jordanian soap opera Al-Mabrukeh, which originally aired on broadcast TV, was selected for transcription. Videos from the broadcast channel Roya which focuses on talk show and panel discussions was selected as well, including the satire program Hakii Jaraid. In addition, web-exclusive videos from the Jordanian production company Kharabeesh were also selected for transcription. Of these four video sources, three hours were selected from each for transcription. Finally, two theater plays totaling approximately three hours of speech that had uploaded to YouTube were selected for transcription. These plays were Alan Fahimtikum, and Ahlan natham 3alami al-jaded, two political plays. The purpose of choosing traditional broadcast content, web-original content and theater scripts was to gain a wider range of topic and genre as to not heavily bias topic and genre inclusion in the final corpus.

Following extraction of the audio, to prepare the audio for transcription, it was processed using the LIUM SpkDiarization toolkit (Meignier and Merlin, 2010), which specializes in speaker segmentation and clustering in audio. Using LIUM, voice activation detection was performed to remove as many non-speech segments (such as music, white noise, silence, or other background noise) as possible. Then, speaker diarization was performed to determine when a new speaker begins a new utterance to lessen the potential of users transcribing speech across multiple speakers for one audio file.

This task was conducted via Crowdflower (crowdflower.com, 2015) (henceforth CF), a crowdsourcing platform which utilizes various worker channels including other microworking sites. CF has many built-in mechanisms to ensure data quality including a researcher-provided gold standard. CF has a robust userbase in the Arab world. In addition, task participation can be limited to users in specific regions, ensuring that participants most closely associated with desired region contributed to the task.

Contributors provided trancripts as follows: after selecting the job, users were presented with five audio files. Each audio file had a play audio button which users could press to listen to the audio an unlimited number of times, and a text box for entering the transcript. Users were directed to listen to the audio and write exactly what they had heard as precisely as possible, to heed the item ID number provided to them, and to avoid using non-Arabic characters. Unbeknownst to the users, one of the five audio files was linked to an audio segment that an expert native speaker had already transcribed as a gold standard. The user's transcript was then compared to the expert transcript. However, because of the variability inherent in non-standard writing of colloquial Arabic, a categorical rejection or acceptance of the user's transcript was not feasible. For this reason, I calculated the difference between the expert transcript and the user transcript and rejected transcripts which did not meet an acceptable overlap threshold, following (Wray et al., 2015). The overlap was calculated using 2.1 where $dist$ refers to Levenshtein edit distance [6] between the transcript and the reference, and $refLen$ refers to the length (in characters) of the expert-provided reference.

[6]Was utilized in CF via JavaScript implementation taken from: https://gist.github.com/andrei-m/982927

$$diff(transcription) = \frac{dist}{refLen} \cdot 100 \tag{2.1}$$

The percentage overlap for an acceptable threshold was 30%, and user-provided transcripts which did not meet this threshold were presented with a warning flag requesting the user to be more precise, and users were prevented from submitting that batch of 5 transcripts until they provided a transcript which met this threshold for that audio file. This method of quality control was shown by Wray et al. (2015) to reduce the instance of spammy transcripts by approximately 6% for crowdsourcing transcriptions of colloquial Arabic speech in broadcast audio.

After completion, users submitted the transcripts and were provided with five more files. The task was entirely self-paced and users could complete as many or as few transcripts as they wanted in batches of 5.

Three separate transcriptions were provided by three different contributors for each audio file, however, only one transcript per file was included in the final corpus. In the end, 226,672 words were transcribed corresponding to approximately 15 hours of audio speech.

2.3.3. Traditional colloquial media

In addition to the aggregation of existing digital resources, web mining for new resources, and transcription of video media described above, several print resources were selected for digitization and inclusion in the corpus. Novels by Jordanian writers Ibrahim Nasrallah, Aiman al-Atoum, Ziad Qasim and Fakhri Qawar, as well as collections of short stories by Fakhri Qawar and poetry by Ibrahim Nasrallah, were digitized. The text was processed using the Qatip Arabic OCR toolkit (Stahlberg and Vogel, 2015). In total, the traditional printed works were 523,890 words.

Finally, after inclusion of traditional media, the makeup of the final corpus is presented in Table 2.6. The transcribed speech includes the media described in section

Type	Total Words
Transcribed speech (including (Maamouri et al., 2007; Appen, 2007)	477,587
Web (including (Almeman and Lee, 2013; Zaidan and Callison-Burch, 2011b) (Mubarak and Darwish, 2014)	14,923,159
Traditional media	523,890
	Total: 15,924,636

Table 2.6: Completed Jordanian Corpus

2.3.2 as well as the Jordanian portions of the Fisher (Maamouri et al., 2007) and Appen (Appen, 2007) mixed Levantine telephone corpora. The web portion of the corpus includes the Jordanian portion of the Levantine tweets collected by Mubarak and Darwish (2014) in addition to tweets harvested as described in section 2.3.1. The web portion also includes portions of the mixed Levantine corpora collected by Zaidan and Callison-Burch (2011b); Almeman et al. (2013) that were classified as Jordanian by the automatic classifier described in section 2.2, in addition to newly harvested material as described in section 2.3.1. As described immediately above, the traditional media includes digitized novels, short stories, and poetry.

2.4. Processing the corpus

To prepare the corpus for usage as a source of lexical statistics used in the following experiment, Experiment 1, as well as the experiments presented in chapter 3 and chapter 4, a process of normalization and morpheme identification had to occur. The purpose of normalization is to reduce inconsistencies typical of colloquial data in order to make the corpus more computationally robust, and the purpose of stemming was to extract the root morphemes of each word in the corpus in order to calculate frequency counts of roots in addition to word frequency counts. (For a complete review of Arabic morphology, see chapter 1.)

To normalize the corpus, a script developed by Darwish et al. (2012) for the purpose of normalization of Twitter data was utilized. This script accomplished two basic tasks: removing non-Arabic characters and/or replacement of non-Arabic characters with their Arabic equivalents, and the reduction of elongated words with repeated characters to the shortest possible token.

An example of the first case is the usage of characters borrowed from Farsi, Urdu, Pashto, or other languages using the Arabic script that are non-standard in written usage of the Arabic language but may be used in informal writing for decorative or stylistic purposes. For example, the Kurdish symbol ۆ is not used in standard usage in written Arabic, but has been attested as a stylistic replacement for و in Twitter data (Darwish et al., 2012). As for reduction of elongated words, the script reduced characters that were repeated for an exaggeration effect (also heavily attested in English web data, i.e. 'woooooooooow!!!!' for 'wow!'.) An example of a text sample and its normalized counterpart is shown in Example 2. As demonstrated, the repeated characters displayed in 2a are reduced in the output shown in 2b.

(2) a. ايششششش؟ شووو هاي؟؟
eʃʃʃ-ʃuuuuu-he
what-what-this
What? What's this?

b. ايش؟ شو هاي؟
eʃ-ʃu-he
what-what-this
What? What's this?

As discussed in detail in chapter 1, an Arabic word is typically analyzed as being composed of two non-contiguous morphemes: the root and the pattern (also known as the binyan). An example of this is the word /daras/ 'to study' which is composed of the consonantal root /drs/ and the interleaved binyan /CaCaC/. In order to perform the root extraction process necessary for the lexical statistics, and the morpheme identification process necessary for calculating binyan productivity statistics, the corpus was processed by utilizing two tools following (El-Defrawy et al., 2015): a light stemmer which extracts the interleaved root+binyan from a word by stripping off its clitics, and a root extractor which separates the root from the word entirely, thus separating it from the binyan.

For a light stemmer, the morphological analyzer and tokenizer MADAMIRA (Pasha et al., 2014) was utilized for its high linguistic accuracy (reported by El-Defrawy et al. (2015) at 91.73%). MADAMIRA is composed of two parts: MADA (Habash et al., 2009) which applies a set of n-gram language models and Support Vector Machines (SVMs) to provide part of speech as well as additional linguistic information for each word in a corpus and by utilizing a robust lexicon as well as the word's context in the specific text. As discussed above, written Arabic exhibits a high degree of homography, meaning trying to attempt to calculate lexical statistics from a raw string without considering the word's context would lead to inadvertently collapsing across categories, for example, counting a noun and verb together because they have the same graphemic form. The second part of the toolkit, AMIRA, also includes a tokenizer and part of speech tagger, but in addition includes a shallow syntactic parser and a named-entity recognition component for the processing of proper nouns. Both parts use the Arabic morphological analyzer SAMA (Maamouri et al., 2010) (the successor of morphological analyzer BAMA (Buckwalter, 2004)) for analysis of MSA words as well as CALIMA-ARZ (Habash et al., 2012b) for Egyptian-trained analysis of colloquial material. The output of MADAMIRA has multiple components, including the tokenized output which separates proclitics such as "and" and "for" as well as enclitics such as the direct object pronouns from the word. MADAMIRA also provides morphological analysis which provides a ranked list of possible analyses as to the underlying lemma of each word in the corpus. From this lemma, the binyan can be derived using simple pattern matching. Although MADAMIRA is not specifically designed to analyze Levantine material, the underlying analysis engine BAMA (Buckwalter, 2004) has been previously shown to provide analysis of up to 71.8% of Levantine (Duh and Kirchhoff, 2005).

After processing the corpus in MADAMIRA for part of speech information and extracting and counting the binyanim of those items determined to be verbs in the corpus, the tokenized corpus was processed by the Java-based Khoja stemmer (Khoja and Garside, 1999), a root extractor that uses pattern-matching and consultation of a built-in lexicon of roots and patterns to determine the root consonants of each word. Words which are generally determined to not have underlying roots, such as /ana:/ "I", are returned as-is without further analysis. Because the potential list of roots is hard-coded in a simple text file read in by the program, it was a trivial task to augment the lexicon with a list of attested Jordanian roots. The Khoja lexicon was manually augmented with roots from various Jordanian and Levantine dictionaries (Elihay, 2004; Azban, 2011; Tiedemann, 2015).

After processing via these toolkits, the corpus contained an amount (approximately 11 percent) of words which were not stemmed as they were unrecognizable words/typos not corrected by normalization, not analyzed by MADAMIRA, or not stemmed by the Khoja stemmer. Future research will entail further analysis and accounting for this material.

2.5. Experiment 1: Validation of corpus as representative sample

In order to determine validity of the corpus as a reasonable source of lexical statistics, a word familiarity task was performed to gather subjective frequency judgments for Jordanian words. This method exploited the correlation between word familiarity judgments provided by native speakers and word frequency counts gleaned from a corpus. It has been shown that the more frequent a word is, the more familiar a participant will rate it in a word familiarity task. This correlation has been demonstrated for a variety of corpora, including a corpus of English monosyllabic words used in past studies (Plaut et al., 1996; Seidenberg and McClelland, 1989; Spieler and Balota, 1997) for which 2,938 word frequency-familiarity counts were found to correlate (Balota et al., 2001). They collected familiarity judgments by presenting words via a web-based platform and directing participants to provide a judgment for each word based on the options in Table 2.7.

Given that subjective familiarities have found to correlate with frequency measurements from established corpora which were known to be robust, it follows then that evaluating a new corpus and determining it represents an authentic language sample can be performed by collecting familiarity judgments and determining if the frequency counts from the newly created corpus correlate. The usage of familiarity-frequency correlation as a methodology for corpus validity has been explored for Khalkha Mongolian (LaCross, 2014) and Maltese (Francom et al., 2010). For Khalkha Mongolian, LaCross (2014) built a corpus by mining web sources such as online newspapers and Wikipedia, then presenting words of varying corpus frequency counts to native speakers and collecting familiarity judgments based on the same familiarity scale developed by Balota et al. (2001). For Maltese, Francom et al. (2010) presented native speakers with words of varying frequency and a sliding scale with "this item is not familiar to me" and "this item is very familiar to me" as the two extremes in order to validate the representativeness of the PsyCoL Maltese Lexical Corpus, a 3.3 million word corpus built largely from online newspapers.

For the current study, Experiment 1, a word familiarity task, was conducted to determine if the assembled corpus was a representative sample of the language local to Jordan. Drawing from this previous work which shows strong correlations between corpus-measured frequency and subjective familiarity (Gernsbacher, 1984; Connine et al., 1990; Francom et al., 2010; Balota et al., 2001; LaCross, 2014), it was hypothesized that high frequency words would be rated as more familiar than low frequency words, and vice versa.

2.5.1. Methodology

Familiarity judgments for a list of Jordanian words were collected from native speakers who rated words on the seven-point scale shown in Table 2.7, following that developed by Balota et al. (2001). Under this system, the participant was presented with a word and asked to select which of the options in Table 2.7 corresponds to that word. For convenience of analysis, a number was assigned post-hoc to each description, with 7 representing the most familiar and 1 being the least familiar. In an attempt to avoid unintentionally priming a participant with a word appearing in both the directions and the stimuli, the directions for this experiment were presented in MSA.

Numeric value	Familiarity Judgment
1	"I never use this word"
2	"I use this word about once a year"
3	"I use this word about once a month"
4	"I use this word about once a week"
5	"I use this word about once every two days"
6	"I use this word about once a day"
7	"I use this word multiple times a day"

Table 2.7: Familiarity Rating Scale

2.5.1.1. Participants

This experiment was performed remotely. Participants were recruited via Crowdflower (CF), and participation in the experiment was limited to residents of Jordan. (See section 2.3.2 above for an overview of this crowdsourcing platform).

In addition to limiting country of participation, further quality control was performed. For every five items that a participant judged, one previously-annotated golden-standard item was included. These included extremely high-frequency words for which a judgment indicating unfamiliarity would be rejected, and nonwords. High familiarity ratings for nonword items were indicative of answers which did not follow proper task procedure, thus all answers from users which exhibited this behavior was discarded.

After exclusions of those who did not conform to quality control standards, 40 participants remained. Participants responded to 1-60 items each, with an average of 12 items per participant and a median of 8 items per participant.

2.5.1.2. Materials

A total of 60 items were recorded by a 23-year-old female life-long resident of Amman who is a native speaker of Jordanian. A female voice was selected to reduce potential ambiguity as male Ammani speakers tend to use phonological features in line with rural/Bedouin pronunciation, such as the usage of /g/ for the Arabic /q/ (Al-Wer and Herin, 2011). Items were recorded on a TASCAM DR-05 Portable Digital Recorder in three repetitions and the clearest token was selected for usage in the experiment.

For this experiment, frequency was treated as a continuous variable, so there was no binning of items into frequency ranges. Items at the lowest frequency end of the range measured a log frequency of -0.58 and the highest log frequency of 7.23. Sample stimuli are shown in Table 2.8. Items were selected across binyanim. A complete list of stimuli is found in appendix A.

Word	Gloss	Binyan	Frequency
masik	'he grabbed'	I (CVCVC)	2.64
iħtaram	'he respected'	VIII (iCtaCaC)	4.17
iħmar	'he reddened'	IX (iCCaC)	4.89
istaɣrab	'he thought strange'	X (istaCCaC)	5.90

Table 2.8: Experiment 1 Stimuli

2.5.1.3. Procedure

Before starting the task, participants were presented with the following directions in MSA: "Words differ in how commonly or frequently they have been encountered. Some words are encountered very frequently, whereas other words are encountered infrequently. The purpose of this study is to rate a list of words with respect to frequency. We believe that your ratings will be important to future studies involving word recognition."

This familiarity task was self-paced. Users of the CF platform could choose to enter and exit the task at anytime without penalty. Participants were presented with a simple audio player via CF's interface that linked to a sound file containing a single word. Participants could play the word as many times as they desired. After listening to the sound file, users were presented with the 7-point familiarity scale in radio buttons. The default selection for the radio buttons was a blank item with no familiarity description, which required participants to perform an action for every sound file. After selecting their chosen familiarity judgment, participants pressed a "submit" button which led them to a new set of audio stimuli awaiting judgment.

2.5.2. Results

As Figure 2.1 illustrates, participants' familiarity ratings aligned with the hypothesized outcome: high frequency words were rated as more familiar than low frequency words. Words with frequency measurements that fell between the two frequency extremes were rated as moderately familiar.

A simple linear regression was performed to predict familiarity judgment based on word frequency. A significant regression equation was found $F(1,453) = 83.49$, $p < .001$, with an R^2 of 0.16.

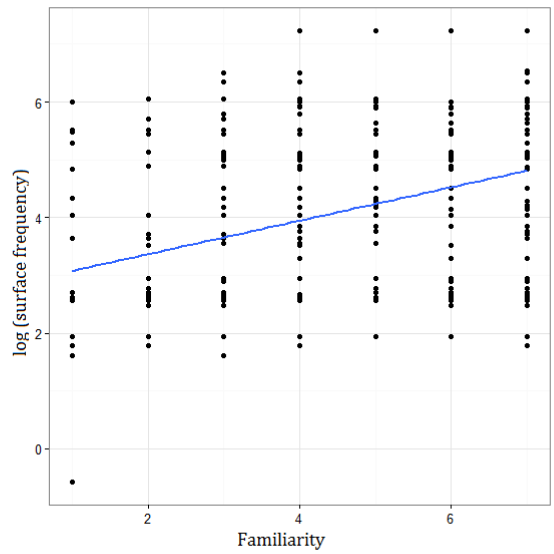

Figure 2.1: Corpus frequency by subjective familiarity

2.5.3. Discussion

Participants' familiarity ratings supported the hypothesis that the corpus was a representative sample of Jordanian by being correlated with the frequency measurements from the corpus. Items which were rated the least familiar were also found to be low frequency words in the corpus, and words which were rated as highly familiar were high frequency.

Given participants' subjective familiarity judgments and their correlation with the frequency counts from the corpus prepared for this dissertation, it can be concluded that the corpus is indeed a representative sample of Jordanian Arabic and a valid source of the lexical statistics utilized for stimuli creation for the lexical decision experiments presented in chapters 3 and 4.

2.6. Conclusions

In this chapter, I described the process of consolidating and creating subsets of existing Levantine corpora into a 5.5 million word Jordanian corpus. Then, I outlined the creation of a 10.5 million word augmentation of Jordanian text from web sources (Twitter, forums), digitization of printed material, and crowd-sourced transcription of audio. I also presented evidence that the complete corpus is representative of Jordanian in the form of a correlation in familiarity judgments gathered in Experiment 1 and frequency measurements from the corpus. Improvements to the corpus in an effort to explore its usage as a balanced and representative sample of Jordanian Arabic are ongoing.

In the following chapters, chapter 3 and chapter 4, I calculate productivity of Jordanian binyanim using the corpus, showing that Binyan I is more productive than Binyan VIII and Binyan X. I also present results of four auditory lexical decision experiments building from these productivity measures as well as word frequencies and morpheme frequencies measured from the corpus.

3. Root frequency in productive verbal paradigms

Semitic words are typically analyzed as being composed of a noncontiguous consonantal root, which is usually tri-consonantal and provides the most salient semantic content, and a noncontiguous pattern (also called binyan) containing vocalic segments and possibly additional consonants and providing grammatical information (as presented in detail in chapter 1). Significant work has been done to demonstrate the psycholinguistic reality of the root morpheme (Boudelaa and Marslen-Wilson, 2004b, 2005, 2011). Further work has been has been done investigating the roles of base and surface frequency of both stems and affixes across various languages (see Baayen et al. (1997), Alegre and Gordon (1999) for Dutch; New et al. (2004) for Finnish; and Taft (1979), Taft (2004) for English) but no study to date has described the morphological structure of Arabic verbs in the lexicon by appealing to evidence from surface frequency of the word and base frequency of the root, which leaves the question of the role affix productivity in lexical access for Arabic verbs unanswered.

As detailed above, Semitic verbs can be organized into verbal paradigms referred to in literature as binyanim (See Chapter 1 for a detailed account of verbal paradigms in Semitic). Some binyanim in Arabic, such as Binyan I of the form CaCaC, are attested as being highly productive in colloquial dialects of Arabic, whereas it has been suggested that others, such as Binyanim III, IV, VIII and X have been lexicalized and are either no longer productive or are less productive than Binyan I and Binyan II verbs (see Watson (2007) for a discussion on Sana'ani and Cairene). The relevance of binyan productivity to the current study is the possibility that verbs which populate lexicalized binyanim may not be as readily decomposable into their constituent morphemes during lexical access as verbs belonging to binyanim which are still productive.

One way of determining the productivity of an affix is to utilize a corpus and count the number of times the affix appears in comparison to the number of "frozen forms" containing the affix that only appear once in the corpus. Words which only appear a single time in a corpus are known as *hapax legomena* (henceforth, hapax). Productivity of an affix can be measured by counting the number of hapaxes which appear in a corpus containing the affix in question and considering their proportion in relation to the total number of hapaxes in that corpus. This hapax-conditioned degree of productivity is formalized here, with ***P**** referring to the calculated productivity measure, *V(1,C,N)* equaling the number of hapaxes with the affix in question and *V(1,C)* being the total number of hapaxes in the corpus (Baayen, 1993; Bauer, 2001).

$$\boldsymbol{P*} = \frac{V(1, C, N)}{V(1, C)}$$

I applied this formula to verbal paradigms to determine their productivity in Jordanian Arabic.. The results are shown in Table 3.1. Measurements were taken from the colloquial Jordanian corpus introduced in chapter 2.

Binyan	Form	***P****
I	CaCaC	0.001395365
II	CaC:aC	0.000794110
III	CaaCaC	0.000397055
IV	ʔiCCaC	0.000884865
V	itCaC:aC	0.001191165
VI	itCaaCaC	0.001247887
VII	nCaCaC	0.000340333
VIII	iCtaCaC	0.000431088
X	staCaCaC	0.000419744

Table 3.1: Hapax-conditioned productivity

As shown in Table 3.1, the hapax-conditioned productivity measures show that Binyan I is more productive than both Binyan VIII and Binyan X, as the hapax-conditioned productivity measure of Binyan I is an order of magnitude larger than these other two binyanim. Implications of these differences in productivity within the framework of lexical access are explored in the current and following chapters. Specifically, it has been found that a complex word's tendency to decompose during the lexical access process is tied to the productivity of the affix it contains (see López-Villaseñor (2012), for Spanish, among others). Words with productive affixes are decomposed during lexical access, as evidenced by the fact words with more frequent base morphemes are recognized more quickly than words with less frequent base morphemes. In contrast, words with unproductive affixes are accessed holistically without decomposition, as evidenced by a lack of the base frequency effect. It is hypothesized that the verbs in the more productive binyan, Binyan I will be decomposed during lexical access and recognition speed will show a base frequency effect. The less productive binyanim, Binyan VIII and Binyan X, will be stored holistically. These binyanim were selected as less productive in favor of other binyanim for several reasons. Although Binyan III is also less productive than Binyan I, it is homophonous with the active participle form and thus it would be difficult to distinguish auditorially when the word is presented in isolation with no further context. Other less productive binyanim are so infrequent that executing an experiment investigating processing speeds of these words would be unfeasible. Because of low frequency, the P* for Binyan IX was not calculated, and therefore is not shown in Table 3.1.

This chapter investigates Binyan I verbs and aims to determine if these verbs are recognized on the basis of their constituent morphemes during lexical access, and if so, what role morpheme and word frequency play in determining how quickly words are recognized. This chapter presents Experiments 2a and 2b, which test the main hypothesis of this dissertation; namely, that both base and surface frequency have a positive effect on recognition speed and facilitate the speed of responses measuring recognition. As outlined above, it was hypothesized that Arabic verbs are fully decomposable into their constituent morphemes during lexical access, meaning that both base and surface frequency affect reaction times (RTs) in a lexical decision task. In this chapter, I present two experiments and discuss their results as evidence that both the base frequency of an Arabic verb and the surface frequency of the word itself affect RTs. I present findings that show that higher base frequency facilitates RTs and results in faster response times, and that higher surface frequencies also facilitate lexical access as evidenced by significantly faster RTs when compared to controls.

3.1. Experiment 2a: Variable base frequency

Experiment 2a specifically targets base frequency by manipulating base frequency as a variable and holding surface frequency constant. This is to determine the contribution of base frequency to reaction time during lexical decision independent of the frequency of the word as a whole.

3.1.1. Methodology

A number of experimental tasks have been used to investigate components of lexical access. For this study, auditory lexical decision was determined to be the most effective way (see chapter 1 for details) to investigate contributions of base frequency of individual morphemes and surface frequency of a whole word to the lexical access process.

In an auditory lexical decision task, participants are presented on each trial with an auditory stimulus and asked to decide, as quickly and as accurately as possible, whether what they hear is a word of their language or not. Lexical decision tasks are especially sensitive to frequency effects when compared with other experimental tasks such as naming and categorization (Balota and Chumbley, 1984). Performing an auditory lexical decision experiment as opposed to a visual lexical decision experiment or a cross-modal experiment which relies on both visual and auditory components circumvents the problem of Arabic orthography's bias to consonantal elements at the exclusion of most vowels which may result in unintended orthography-related effects arising in an experiment. In addition, as outlined in chapter 1, because colloquial

Arabic has only recently been widely written down, a visual task involving judgments on colloquial words written in isolation may be too unfamiliar a task for participants.

3.1.1.1. Participants

For Experiment 2a, 66 speakers of Jordanian were recruited at The University of Jordan in Amman, Jordan. Because of the absence of an institution-wide recruitment system for experiments of this nature, participants were recruited exclusively by word of mouth.

Speakers were determined as Jordanian-dominant by both self-report and completion of a language history survey which collected demographic info including language history (as found in Appendix B). Data from speakers who were determined to be native speakers of another dialect of Arabic or another language were not retained for analysis. All participants reported normal hearing and normal or corrected-to-normal vision. Furthermore, only speakers who maintained at least 70% accuracy throughout the experiment had their data retained for analysis.

After these exclusions, Experiment 2a included a final total of 60 participants of which 32 were female and 28 were male. The ages of participants ranged from 18-31, with an average age of 21 and a median age of 21. An additional 5 participants were excluded for being a native speaker of a dialect other than the target dialect, and 1 participant was excluded for completing the experiment with less than 70% accuracy.

3.1.1.2. Materials

The test items used in Experiment 2a were 60 verbs from Binyan I which varied in the frequency of their root morphemes (base frequency). Surface frequency was controlled across items. All items had a low surface log parts per million frequency ranging from 0.124-0.784 with a mean value of 0.436.[1] Both base and surface frequency were

[1]Given the low frequencies and the ubiquity of code-switching, speakers may consider less frequently encountered words as more MSA than colloquial. However, pilot testing of items indicated

considered continuous variables for this experiment. The stimuli distribution is shown in Figure 3.1 which plots test items in base and surface frequency. Sample stimuli from each of these types is shown in table 3.2 at three slices of the frequency continuums: low base frequency/low surface frequency; mid base frequency/low surface frequency; and high base frequency/low surface frequency. A complete list of test items for Experiment 2a is shown in Appendix C.

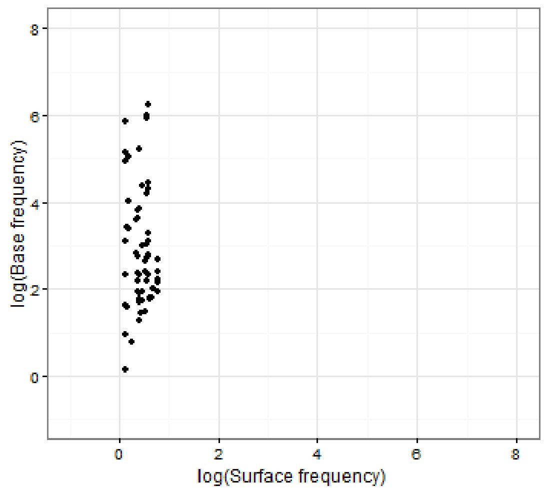

Figure 3.1: Experiment 2a stimuli base x surface frequency

	low base/ low surface	mid base/ low surface	high base/ low surface
Word	hamaz	ʔatˤam	ʕaraʔ
Gloss	"he backbit"	"he cut"	"he sweat"
Root	h-m-z	ʔ-tˤ-m	ʕ-r-ʔ
Base frequency (log)	2.22	2.70	4.00
Surface frequency (log)	0.77	0.56	0.18

Table 3.2: Experiment 2a sample stimuli from three points in frequency continuum [2]

In addition to these test items, 60 distractors were included in the experiment. These distractors were intended to ensure that participants did not develop a strategy

they were readily identified as real words of Arabic.

for responding to stimuli that obscured the desired effect, i.e. the roles of base and surface frequency and their relationship to the speed at which participants recognized words. Distractors were real words of Jordanian Binyanim III, IV, V, VI, VII, VIII, X, as well as frozen Jordanian-MSA cognates from Binyan IX of the form iCCaC. No Binyan I items were included in the distractor set.

As a lexical decision task requires nonwords to necessitate the required negative responses, 120 nonwords were included in the experiment. This is an equal number of nonwords as real words, ensuring that participants would be required to provide a 'No' response exactly as many times as a 'Yes' response. Nonwords were created by generating a list of possible roots that were unattested in Jordanian and neighboring dialects of the Levant (Syrian, Palestinian, Lebanese) but did not violate phonotactic constraints of the language (as in Ussishkin et al. (2015)). Then, the possible-but-not-attested roots were combined with existing verbal binyanim to ensure diversity in the nonword pool just as in the distractor pool. Nonwords had their nonexistence confirmed by 1) searching for them in lexicons including Elihay (2004) 2) searching for them in the corpus 3) obtaining native speaker judgments from an informant regarding their lexicality.

Items were recorded by a 23-year-old female native speaker of Jordanian who is a life-long Ammani resident using a TASCAM DR-05 Portable Digital Recorder. A female speaker was selected as speakers in Jordan tend to have gendered phonemes depending on region/etc. A male speaker in Amman may use features that correspond with a more rural/Bedouin pronunciation, whereas female speakers in rural areas and in Amman have distinct sound inventories. For example, male speakers in both rural Jordan utilize /g/ as an alternate for the standard Arabic /q/, whereas rural Jordanian women use /g/ and Ammani women use /ʔ/ (Al-Wer and Herin, 2011). Therefore, using the recordings of a female speaker as opposed to a male speaker helped code the speech as unambiguously Ammani to the listener.

The speaker was directed to record each word three times to create a carrier frame for a middle token with neutral prosodic intonation. The middle token was selected for use as a stimulus except in cases where the middle token was found to be corrupted by the presence of audible non-linguistic noises. The speaker was unaware of the conditions of the experiment upon the time of recording to help prevent affected pronunciation and ensure the most neutral pronunciation possible.

3.1.1.3. Procedure

Upon completion of the language history and demographic survey, participants were seated in front of a computer screen on a portable ASUS computer in a quiet room. Participants were directed to wear Sennheiser HD 598 Over-Ear Headphones through which the auditory stimuli was presented. Participants were also directed to both listen to the experimenter explain the experimental procedure as well as read the directions on the computer screen. The experiment was presented to participants using DMDX (Forster and Forster, 2003), which also collected data on RTs from stimulus onset and accuracy. نعم ('Yes') and لا ('No') labels marked the right and left shift keys on the keyboard from which responses were recorded. Before the random stimuli presentation part of the task, participants completed a 12 item practice session from which data was not retained for analysis. This was performed to ensure that participants understood the task and to provide a period of acclimation such that participants could become familiarized with the experiment procedure. No target items from the experiment also occurred in the practice session. Following the practice session, items were presented in a random order with no constraints that was randomized for each participant. To keep participant focus engaged, accuracy feedback was displayed visually on the screen immediately following each lexical decision (صحيح 'Correct', مش صحيح 'Incorrect'). Responses timed out at 3000 ms to ensure that participants answered as quickly as possible, and participants were explicitly encouraged to answer as quickly as possible while still maintaining accuracy.

3.1.2. Results

Sample Reaction Times (RTs) and accuracy were recorded for each item by DMDX. Ten items (ħaraʃ, balatˤ, ħalal, ʕasˤab, ʃaʕal, nisum, ɣalaf, taledʒ, ʔazam, ʕaʃam) were excluded from analysis, as they had been incorrectly coded as Binyan I verbs but were in fact Binyan II verbs, or inter-speaker variation regarding the presence of a medial geminate on the verb (ʔazam and ʔaz:am, for example) was discovered post-hoc and would place the verb in either Binyan I or Binyan II depending on the speaker.

Incorrect responses were discarded, as well as responses to distractor items and nonwords. Test items which were answered correctly were analyzed using statistical tests run using R (R-Core-Team, 2014). A linear mixed effects analysis (using (Pinheiro et al., 2016)) was performed to determine the relationship between base and surface frequencies and RTs. This method also accommodates multiple random effects (in this case, subjects and items), and analyzes RTs individually instead of combining RTs for items and RTs for subjects (Baayen et al., 2008; Winter, 2013).

Test items with RTs of less than 300 ms were excluded from analysis, and RTs more than 3 standard deviations from each participant's mean were excluded from analysis. Base frequency and surface frequency were treated as continuous fixed effects, and intercepts for subjects and items were treated as maximal random effects, following Barr et al. (2013). A fixed effect of item duration in ms, and a fixed effect of trial number, as well as by-subject and by-item random slopes and random slopes for trial number by item and by subject were also included in the model.

Typically in mixed effects models, p-values are calculated on the upper bound of the degrees of freedom as there is no clear establishment on the estimation of degrees of freedom (Baayen et al., 2008; Bates, 2005). P-values were obtained from the mixed effects model using the t-distribution, and thus significance was interpreted from the t value.

Raw means and standard deviations are shown in table 3.3. RT distribution is shown in Figure 3.2.

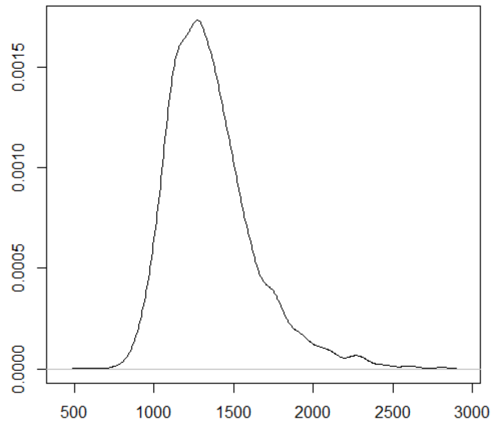

Figure 3.2: Raw Reaction Times for Experiment 2a

As shown in Figure 3.2, RTs were skewed. To account for the positive skew in RTs, analyses were run on log(RT), however results did not vary when raw RTs were analyzed. There was a significant main effect of base frequency on RT $t = -2.08, p < .04$. Specifically, as base frequency increased, RTs were shorter, indicating that participants responded more quickly to words with higher base frequency. There was also an effect of item duration $t = 3.81, p < .005$. There was no significant interaction between base and surface frequencies. $t = -.27, p = .78$.

	low surface/ low base	low surface/ mid base	low surface/ high base
Mean RT	1476.16	1364.19	1301.64
Standard Error	±12.57	±10.12	±9.97
Confidence Interval (95%)	±24.68	±19.48	±19.33
Total observations discarded	658 (21%)	436(11%)	548(13.8%)

Table 3.3: Experiment 2a Means

3.1.3. Discussion

The results of the analyses of Experiment 2a indicate that base frequency of the root morpheme predicted how quickly speakers were able to recognize the word as evidenced by a significant main effect of base frequency and the facilitation in RTs for words with more frequent bases.

Results of Experiment 2a support the hypothesis that the base frequency of a word affects RT, and furthermore support the hypothesis that Arabic verbs are decomposable into their constituent morphemes during lexical access. As surface frequency did not vary across items, it was expected that surface frequency would not be shown to affect RTs in Experiment 2a, and indeed the results for Experiment 2a show that surface frequency was not found to have a significant effect. Crucially, as base frequency varied across items and it was hypothesized that base frequency plays a role in the lexical decision process, it was expected that base frequency would be the sole predicting factor in determining RT, and this was supported in the results.

Given that the decomposability of Arabic verbs and the psycholinguistic saliency of the root has been reliably replicable across modalities (such as in Boudelaa and Marslen-Wilson (2004b, 2005, 2011) among others), it should come as no surprise that the current results reflect support for a model that incorporates decomposability of the word such that the root is available in lexical access. Indeed, the results of Experiment 2a support these models by showing that the base frequency of a word affects the speed at which a speaker can recognize that word when exposed to it auditorily. In addition to previous studies on base morphemes in Arabic, these results are consistent with work on base morphemes in English (Baayen et al., 1997; Ford et al., 2010) Italian (Burani et al., 1984), and Dutch (Baayen et al., 1997). Evidence from Italian (Burani et al., 1984) showed that words are processed as a function of their base frequency. For English, (Ford et al., 2010) has shown that base morpheme frequency facilitated RTs for words which were productively suffixed. Further work on

English and Dutch has shown the influence of base frequency on RTs was dependent on which derivational affix the word contained (Baayen et al., 1997). As further explored in the following chapter, this dependence holds true for Arabic as well, as verbs which have less productive Binyanim affixes are not processed with respect to their base frequencies.

The results of the current experiment are not consistent with a case of reverse base frequency effect (Taft, 2004) in which low surface frequency words did not exhibit a facilitatory effect of base frequency. On the contrary, the higher the base frequency, the longer the RT. This is interpreted as evidence for an obligatory decomposition model in which a word is automatically broken down into its constituent morphemes, and evidence against a hybrid model, as an obligatory decomposition model would account for the fact that words with high base frequency would suffer an inhibitory effect during recombination of affix and base. However, the results from Experiment 2a support a hybrid model as there is no inhibitory effect of a high base morpheme.

This experiment has shown that base frequency is undoubtedly a factor in predicting RTs for this set of words as the set of words did not differ in their surface frequency because surface frequency was held constant across items. If base frequency of the root morpheme had not affected the speed at which a speaker recognizes the word, and if the word had not been decomposable into its constituent morphemes as suggested by Taft (2004) but had merely been accessed in whole form as suggested by Tyler et al. (1988); Gonnerman et al. (2007), among others, then the expected result would be no significant difference in RTs as surface frequency did not vary.

However, the results from Experiment 2a alone cannot be used to interpret the role of surface frequency in lexical access. To determine if the surface frequency of the whole word significantly affects RTs, another experiment, Experiment 2b, was performed.

3.2. Experiment 2b: Variable surface frequency

Given that results from Experiment 2a demonstrated that the base frequency of the root morpheme significantly affected RTs in a lexical decision task, a companion experiment was necessary to determine if surface frequencies played a similar role. For Experiment 2b, base frequency was held constant across the test items of the stimuli, and surface frequency of the whole word varied.

3.2.1. Methodology

The same methods as Experiment 2a were utilized for Experiment 2b.

3.2.1.1. Participants

Sixty-nine participants for Experiment 2b were recruited in the same manner as Experiment 2a, and also were recruited at the University of Jordan. Exclusion criteria were the same for both Experiment 2a and 2b. A total of 63 participants were included in the final analysis of the experiment, of which 49 were female and 14 were male. The ages of participants ranged from 18-22, with an average age of 20 and a median age of 20. Data from 1 additional participant was excluded from the experiment for being a native speaker of a non-target dialect, and 5 participants were excluded for completing the experiment with an accuracy of less than 70%.

3.2.1.2. Materials

The materials used for Experiment 2b were similar to 2a in that they were drawn from the corpus described in chapter 2 and vetted by native speaker judgments, and recorded by the same native speaker who recorded the stimuli for Experiment 2a using the same recording protocols to ensure recording quality and consistency. However, in contrast with the stiumli used for Experiment 2a, the verbs used as stimuli for this experiment did not differ on the basis of the base frequency of their root morphemes, but rather on the surface frequency of the words themselves.

Experiment 2b included 60 verbs from Binyan I which varied in their surface frequency. Base frequency was held in the low/mid range across items, from 2.92-3.84 log parts per million, with an average value of 3.29.[3] Stimuli are shown plotted surface x base frequency in Figure 3.3. Base and surface frequency were continuous variables. Sample items from these conditions are shown in table 3.4, at three parts in the frequency continuums: low base frequency/low surface frequency; low base frequency/mid surface frequency; low base frequency/high surface frequency. The full stimuli list for Experiment 2b can be found in Appendix D.

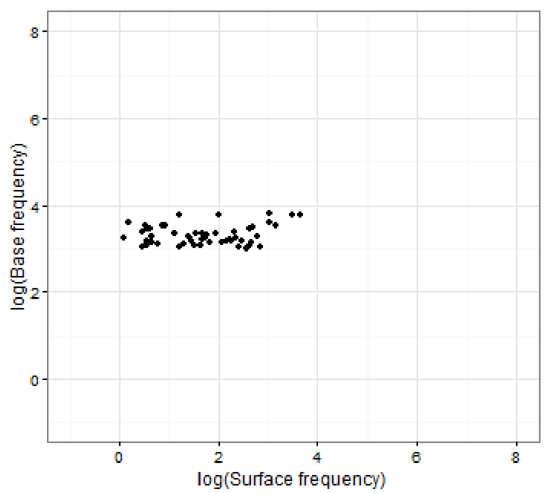

Figure 3.3: Experiment 2b stimuli base x surface frequency

The same distractor items (60) and nonwords (120) from Experiment 2a were also utilized in Experiment 2b to ensure participants did not develop a strategy which obscured the desired effect (i.e. effect of surface frequency), and to enable participants to provide a negative response to half of the task items.

[3]Note that a low value for base frequency is not in the same range as a low value for surface frequency, given that all base frequencies are higher as it is impossible for the surface frequency for a given word to be higher than the frequency of its base morpheme.

	low-mid base/ low surface	low-mid base/ mid surface	low-mid base/ high surface
Word	kabas	zamar	ramaʃ
Gloss	"he pressed"	"he honked"	"it flickered"
Root	k-b-s	z-m-r	r-m-ʃ
Base frequency (log)	3.10	3.17	3.52
Surface frequency (log)	0.77	2.24	3.14

Table 3.4: Experiment 2b sample stimuli from three points in frequency continuum

3.2.1.3. Procedure

The procedure implemented throughout the execution of Experiment 2b was identical to that of Experiment 2a.

3.2.2. Results

Statistical models for the analysis of results of Experiment 2b were constructed in the same manner as those used to analyze the results of Experiment 2a. Twelve items (nasim, ħasar, sumim, fataʃ, nadʒid, ʕasar, ħatˤab, rafaʃ, xasˤal, maʃat, faħam) were excluded from analysis, as they had been incorrectly coded as Binyan I verbs but were in fact Binyan II verbs, or speaker variation placing the verb in either Binyan I or Binyan II was discovered post-hoc.

Raw means and standard deviations are shown for item type in table 3.5. Test items with RTs of less than 300 ms were excluded from analysis, and RTs more than 3 standard deviations from each participant's mean were excluded from analysis. There was a significant main effect of surface frequency on RT $t = -2.01, p < .05$. There was an additional effect of target duration $t = 4.16, p < .0001$. There was no significant interaction between base and surface frequencies. $t = -.33, p = .74$.

	low-mid base/ low surface	low-mid base/ mid surface	low-mid base/ high surface
Mean RT	1465.9	1417.3	1374.87
Standard Error	±12.58	±8.87	±11.11
Confidence Interval (95%)	±24.70	±16.39	±21.81
Total observations discarded	591 (14.3%)	234 (6%)	578 (14%)

Table 3.5: Experiment 2b Means

3.2.3. Discussion

The results of Experiment 2b indicate that surface frequency of the word affects the speed at which speakers recognized the words, and this is demonstrated by the significant effect of surface frequency and the facilitatory effects of frequency: the more frequent a word is, the more quickly is was recognized as evidenced by the significantly faster RT when compared to less frequent words. Note that base frequency did not significantly affect RTs, and this was to be expected as the frequency of the root morphemes in each word was kept constant in order to isolate the effect of surface frequency without the attested effects of base frequency as seen in Experiment 2a. These results are consistent with results from English (Taft, 1979; Burani et al., 1984) which show that when base frequency is held constant, multimorphemic words with high surface frequency are recognized faster. Furthermore, these results align with evidence from Italian (Burani et al., 1984) which showed that words are processed as a function of their surface frequency, as well as evidence from French (Colé et al., 1989) which showed that some multimorphemic words (those which are affixed as opposed to suffixed) are processed whole-word.

These results support the hypothesis that surface frequency in addition to base frequency has a role during the auditory lexical access process and that models which posit a component entailing whole word facilitated access such as those in Taft (1979); Burani et al. (1984) are supported.

3.3. General Discussion of Experiments 2a and 2b

Collectively, the results from Experiments 2a and 2b support the hypothesis that the speed at which words are recognized is affected by both surface and base frequency. Evidence supporting the important role of base frequency (that is, of root morphemes) in facilitating lexical access and resulting in significantly faster RTs in a lexical decision experiment (Experiment 2a). Similar evidence which supports the facilitatory effect of surface frequency on lexical decision was found in the results of Experiment 2b.

Each of the three types of models reviewed in chapter 1 offers an interpretation of these results. In a holistic model, the surface frequency of the word affects the speed of recognition, and in an obligatory decomposition model, the frequency of a word's constituent morphemes also affect speed of recognition. Under a hybrid model, words are recognized via multiple pathways and can be recognized holistically or as a function of the morphemes they contain. The results of the experiments presented in this chapter show evidence of decomposition into constituent morphemes. However, the results also support a hybrid model of lexical access in which words are recognized in two ways: by their surface forms, and by the morphemes which comprise them.

However, recall that both Experiments 2a and 2b exclusively focused on verbs which populate the paradigm of Binyan I, which is a productive paradigm. The question then arises: do verbs which belong to nonproductive verbal paradigms behave in the same manner during lexical access? As the following chapter, chapter 4, discusses in detail by the way of two additional experiments, words which contain less productive verbal affixes are not fully decomposable into their constituent morphemes during lexical access in contrast with the productive verbal paradigm explored here through Experiments 2a and 2b. The following chapter presents and discusses results from additional experiments aiming to determine the role of affix productivity in support of a hybrid model of lexical access by examining how speakers process words from a less productive binyan.

4. Root frequency in less productive verbal paradigms

This chapter presents Experiments 3a and 3b. Both experiments are auditory lexical decision tasks, and both experiments contained verbs which varied based on both their surface and base frequencies. Recall that Experiments 2a and 2b contained verbs from the verbal paradigm Binyan I, a highly productive paradigm in Arabic of the form /CaCaC/. Experiment 3a focused on a particular verbal paradigm, which I have shown is less productive according to its hapax-conditioned degree of productivity: Binyan VIII of the form /iCtaCaC/ (see chapter 3 for details on hapax-conditioned productivity). This Binyan is less populated than Binyan I in addition to being less productive. Experiment 3b is identical to Experiment 3a except that the experiment uses verbs from Binyan X of the form /staCCaC/. Much like Binyan VIII, Binyan X is less-populated than Binyan I and is less productive in contrast with the productive Binyan I.

Productivity and its relationship to decomposability during lexical access with respect to Arabic verbs was explored by Boudelaa and Marslen-Wilson (2011). It was found in a visual masked priming study and a cross-modal priming study that decomposability during lexical access is driven entirely by the productivity of the root, and that productivity of the binyan was irrelevant for nouns. However, evidence from other languages has shown that the productivity of the derivational affix is what determines decomposition during lexical access. In English, Ford et al. (2010) provided evidence that base morpheme frequency facilitated RTs for words which were productively suffixed in a visual lexical decision task. Words with unproductive suffixes did not exhibit an effect of base frequency. Furthermore, evidence from Spanish (López-Villaseñor, 2012) showed that the presence of a base morpheme effect during auditory lexical decision depends on the productivity of the derivational affixes. Productive affixes trigger decomposition as evidenced by the existence of a base frequency effect and unproductive affixes result in the absence of this effect, meaning words are processed whole-form.

In this chapter I present evidence that in an auditory lexical decision task, only surface frequency affects the speed at which speakers recognize verbs from Binyan VIII and Binyan X, in contrast with the results presented in the previous chapter which demonstrated roles for both surface and base frequency in predicting speed of word recognition. This suggests that binyan productivity does indeed play a role at some level of lexical access for Arabic speakers. If lexical access across modalities were driven by root productivity, then words would be fully decomposable into their constituent morphemes and recognized by those morphemes regardless of binyan, which would result in RTs to words being affected by both surface and base frequency, as was shown in Experiments 2a and 2b. On the contrary, I hypothesize that binyan productivity will determine if a verb is recognized by its constituent morphemes during lexical access, and I test this hypothesis by varying base and surface frequency for Binyan VIII and Binyan X, which I have shown are less productive than Binyan I. As demonstrated by Baayen et al. (2007); Bertram et al. (2000b); López-Villaseñor (2012); Taft and Nguyen-Hoan (2010), derivational affix productivity determines whether words are recognized holistically, meaning as a function of their surface frequencies, or whether words are recognized by their constituent morphemes, meaning base morpheme frequency would predict recognition speed. It is therefore hypothesized that Binyan VIII and Binyan X words will be recognized at speeds dependent on their surface frequencies, and that base frequency will not show a significant effect.

Because the less productive paradigms are less populated than Binyan I (estimates from the corpus introduced in chapter 2 show that Binyan I has 1,789 unique types in contrast with 481 for Binyan VIII and 383 for Binyan X), Binyanim VIII and X do not contain enough verbs to provide stimuli for two experiments exactly paralleling Experiments 2a and 2b. Recall that Experiments 2a and 2b were constructed such that one experiment (Experiment 2a) contained Binyan I verbs which varied with respect to their base frequency, but which did not vary with regards to their surface frequency, and the companion experiment (Experiment 2b) contained additional Binyan I verbs which varied by surface frequency but did not vary with respect to surface frequency. However, given that Binyan I is very populous and Binyan VIII and Binyan X are not, Experiments 3a and 3b are constructed so that both base and surface frequencies vary across stimuli simultaneously, but independently. Thus, Experiment 3a contains test items for which the roots are relatively high frequency in relation to the frequency of the words themselves, and vice versa. The items used for each experiment are described in detail in the Materials sections corresponding to each experiment below.

4.1. Experiment 3a: Binyan VIII Verbs of Varying Base and Surface Frequency

Experiment 3a sought to uncover the roles of both base frequency and surface frequency in affecting the speed of word recognition. Both frequency types were manipulated independently in order to assess the effects of each. This experiment focused on verbs in an unproductive verbal paradigm, Binyan VIII. Verbs in this paradigm are characterized by the pattern iCtaCaC where C indicates a slot in where the root consonants interleave during word formation.

4.1.1. Methodology

The methodology for this experiment was similar in form to Experiments 2a and 2b, as it was also an auditory lexical decision task. There were key differences in stimuli, including the fact that both variables (surface frequency, base frequency) varied across stimuli. In addition, the set of distractors used included more Binyan I items and did not include Binyan VIII items. These differences are noted in detail in the Materials section.

4.1.1.1. Participants

For Experiment 3a, 62 speakers of Jordanian were recruited at The University of Jordan in Amman, Jordan, and Qasid Institute in Amman, Jordan. As in Experiment set 2, participants were recruited exclusively by word of mouth. Speakers were determined as Jordanian-dominant by both completion of a language history survey and self-report. As in Experiment 1, data from speakers who did not finish the experiment with at least 70% accuracy or those who were determined to be native speakers of another dialect of Arabic or another language were not retained for analysis.

All participants in Experiment 3a reported normal hearing and normal or corrected-to-normal vision. Experiment 3a included a final total of 60 participants of which 38 were female and 22 were male. The ages of participants ranged from 18-31, with an average age of 22 and a median age of 21. An additional 2 participants were excluded for being native speakers of another dialect. No participants who participated in Experiments 2a and 2b were included in Experiment 3a.

4.1.1.2. Materials

Recall that due to the fact that Binyan VIII is a less populated binyan and less productive binyan than Binyan I, there were not enough items to warrant two experiments, one in which the items varied by surface frequency and another in which the items varied by base frequency. Instead, 60 items which varied in both base and surface frequency were selected for inclusion. Both base and surface frequency were continuous variables. Sample stimuli are shown in table 4.1, at four points in the frequency continuums: mid base frequency/low surface frequency; high base frequency/low surface frequency; mid base frequency/high surface frequency; high base frequency/high surface frequency. These four types covered the whole frequency range for Binyan VIII verbs in the Jordanian dialect. The full stimuli are plotted by base and surface frequency in Figure 4.1.

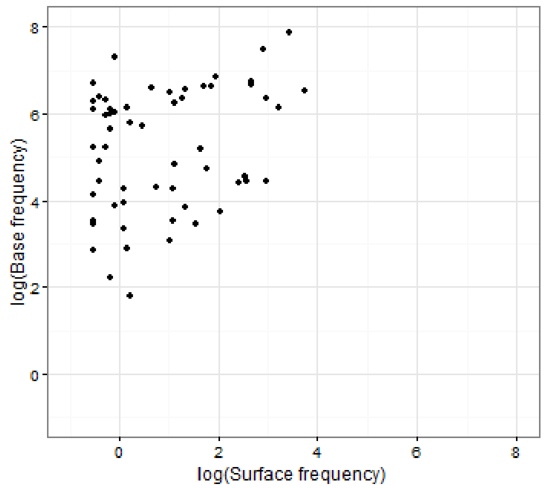

Figure 4.1: Experiment 3a stimuli base x surface frequency

	mid base/ low surface	high base/ low surface	mid base/ high surface	high base/ high surface
Word	intaʃal	iftaʕal	ixtalatˤ	ijtamaʕ
Gloss	"he snatched"	"he fabricated"	"it mixed"	"he accumulated"
Root	n-ʃ-l	f-ʕ-l	x-l-tˤ	dʒ-m-ʕ
Base freq(log)	2.87	5.71	4.41	7.47
Surface freq(log)	0.15	0.45	2.41	2.91

Table 4.1: Experiment 3a sample stimuli from four points in frequency continuum

In addition to the 60 test items, 60 distractors were included to prevent participants from developing a strategy to respond to test items that would obscure the effect in question (roles of base and surface frequency in affecting RT.) These distractors were borrowed from Experiments 2a and 2b. However, the Binyan VIII verbs from that distractor list were replaced with Binyan I verbs. Finally, 120 nonword items were included to ensure that participants answered 'No' to the lexical decision task as often as they answered 'Yes.' The nonword items were identical to those used in Experiment 2a and Experiment 2b.

4.1.1.3. Procedure

The procedure used for Experiment 3a was identical procedure for Experiment set 2.

4.1.2. Results

Statistical models for the analysis of results of Experiment 3a were constructed in the same manner as those used to analyze the results of Experiments 2a and 2b. Test items with RTs of less than 300 ms were excluded from analysis, as well as items which were outside 3 standard deviations from each participant's mean.

Raw means and standard deviations are shown in table 4.2. There was a significant main effect of surface frequency on RT $t = -2.69$, $p < .01$, and no significant effect of base frequency $t = -.51$, $p = .60$. There was an additional effect of item duration $t = 4.14$, $p < .0001$. There was no significant interaction between base and surface frequencies. $t = -.66$, $p = .50$.

	mid-high base/ low surface	very high base/ low surface	mid-high base/ high surface	very high base/ high surface
Mean RT	1374.74	1319.95	1322.90	1298.10
Stnd Error	±10.92	±8.899	±10.57	±10.35
CI (95%)	±21.44	±17.46	±20.75	±20.32
Discarded	140 (4%)	72(1.9%)	104 (2.8%)	62 (1.6%)

Table 4.2: Experiment 3a Means

4.1.3. Discussion

Results of Experiment 3a suggest that the surface frequency of a Binyan VIII word is a predictor of how quickly speakers were able to recognize that word, as evidenced by the significant main effect of surface frequency on RT. There is no significant effect of base frequency on RT, which is consistent with the hypothesis that these verbs are not decomposed into their constituent morphemes during lexical access. These results contrast with those found for Experiments 2a and 2b which showed that both surface frequency and base frequency affected RT.

Although results from Experiment 2a and 2b provide evidence that the speed at which Binyan I words are recognized is affected by both the base frequency of the root morpheme and the surface frequency of the word itself, recall it was hypothesized that unproductive binyanim are populated with Arabic verbs which are not fully recognized by their constituent morphemes in contrast with the verbs from productive binyanim such as Binyan I. Results from Experiment 3a support the hypothesis that Arabic verbs belonging to the unproductive verbal paradigm Binyan VIII are recognized by their surface forms, and not by their base morphemes. This indicates that they are not decomposable into their constituent morphemes all throughout the lexical access process. In the following section, I present evidence through the results of Experiment 3b that an additional unproductive paradigm, Binyan X, is also not decomposable throughout the lexical access process. Just as results from Experiment 3a have shown that surface frequency is a predictor of RT, Experiment 3b presents additional evidence by showing that surface frequency predicts RT in a lexical decision task containing Binyan X verbs as well.

4.2. Experiment 3b: Binyan X Verbs of Varying Base and Surface Frequency

4.2.1. Methodology

The methodology for this experiment was similar in form to Experiments 2a and 2b and Experiment 3a. This experiment was also an auditory lexical decision task. Much like Experiment 3a, the set of distractor items used for this experiment used Binyan I items and did not include Binyan X items.

4.2.1.1. Participants

Sixty-three participants were recruited at The University of Jordan and Qasid Institute. Recruitment procedures were identical to those outlined in Experiment set 2 and Experiment 3a, and exclusion criteria also remained the same. All participants in Experiment 3b were determined to have normal hearing and normal or corrected-to normal vision. A total of 60 speakers participated in Experiment 3b, with an additional 3 speakers being excluded for failing to meet the 70% accuracy threshold. Of those participants included in the final analysis, this included 51 females and 9 males, with ages ranging from 18-40 with an average age of 22 and median age of 21. No participants who participated in Experiments 2a, 2b or 3a were included in Experiment 3b.

4.2.1.2. Materials

Recall that Binyan X is less populated and less productive than Binyan I. Thus, there were not enough items to spread across two experiments to mirror the construction of Experiment 2a and 2b. For Experiment 3b, 60 items which varied across both the base frequency of the root morpheme and the surface frequency of the whole word. Both base and surface frequency are continuous variables in Experiment 3b. Sample stimuli are shown in table 4.3, at four points of the two frequency continuums: mid base frequency/low surface frequency; high base frequency/low surface frequency; mid base frequency/high surface frequency; high base frequency/high surface frequency. These four types covered the whole frequency range for Binyan X verbs in the Jordanian dialect. The full list of stimuli is presented in Appendix F. Stimuli are shown in a base x surface frequency plot in Figure 4.2.

In addition to the 60 test items, 60 distractor items were included. The list was substantially the same as the distractor list used in Experiments 2a and 2b. However, the Binyan X items were removed from the list and replaced with Binyan I

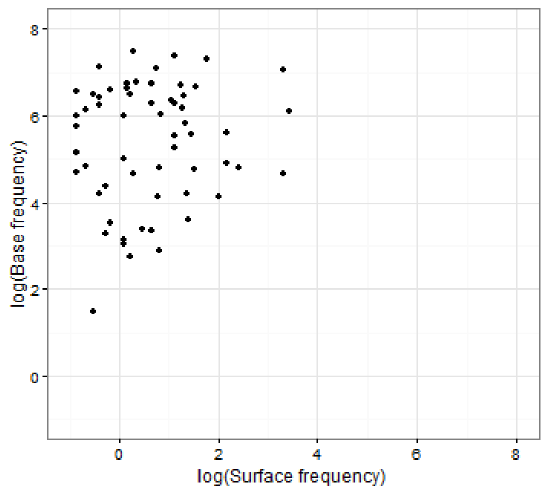

Figure 4.2: Experiment 3b stimuli base x surface frequency

	mid base/ low surface	high base/ low surface	mid base/ high surface	high base/ high surface
Word	stafraɣ	stakmal	staɣfar	staɣrab
Gloss	“he vomited”	“he completed”	“he apologized”	“he found strange”
Root	f-r-ɣ	k-m-l	ɣ-f-r	ɣ-r-b
Base freq(log)	3.53	6.25	4.66	6.09
Surface freq(log)	-0.18	-0.40	3.29	3.41

Table 4.3: Experiment 3b sample stimuli from four points in frequency continuum

items. Finally, as in Experiments 2a, 2b, and 3a, 120 nonwords were also included in Experiment 3b to necessitate a negative response to half of the items in the task.

4.2.1.3. Procedure

The procedure used for Experiment 3b was identical to that used in Experiment set 2 and Experiment 3a.

4.2.2. Results

Statistical models for the analysis of results of Experiment 3b were constructed in the same manner as those used to analyze the results of Experiments 2a, 2b, 3a. Test items with RTs of less than 300 ms were excluded from analysis as well as RTs not within 3 standard deviations for each participant's mean.

Raw means and standard deviations are shown in table 4.4. There was a significant main effect of surface frequency on RT $t = -2.00$, $p < .05$, no significant effect of base frequency $t = -0.90$, $p = 0.36$. There was also an effect of trial number in this experiment $t = 2.90$, $p < .005$. As trial number went up, participants took longer to respond. This is indicative of a fatigue effect. There was no significant interaction between base and surface frequencies. $t = -.68$, $p = .49$.

	mid-high base/ low surface	very high base/ low surface	mid-high base/ high surface	very high base/ high surface
Mean RTs	1383.96	1378.64	1356.89	1332.05
Stnd Error	±10.09	±9.56	±9.59	±8.72
CI (95%)	±19.22	±18.60	±17.99	±16.87
Discarded	166 (4.3%)	139 (3.6%)	151 (3.9%)	183 (4.8%)

Table 4.4: Experiment 3b Means

4.2.3. Discussion

Results of Experiment 3b suggest that, in contrast with the results presented in Chapter 3 for Experiments 2a and 2b, that the surface frequency of a word is a significant predictor of how quickly speakers were able to recognize said word, and base frequency is not. Evidence of this was shown by the significant main effect of surface frequency and the facilitation in RTs for higher frequencies.

Results of Experiment 3b support the hypothesis that productivity of the binyan determines if the word will be stored holistically, as evidenced by the surface frequency effect. The results also are consistent with a lack of decomposition as there was not found to be a base frequency effect. These results suggest that not all Arabic verbs are fully decomposable into their constituent morphemes during lexical access, and that the determiner of this decomposability is the productivity of the binyan. Across all conditions regardless of the relation between surface frequency and base frequency (such as if base frequency was relatively high compared to surface frequency or vice versa), base frequency was irrelevant and did not significantly affect RTs. If these words had been fully decomposable into their constituent morphemes, it would be expected that RTs would be facilitated for words which contained higher frequency root morphemes. However, as shown in the results, this was not shown to be the case as there was so significant effect of base frequency.

4.3. General discussion of Experiments 3a and 3b

Together, the results of Experiments 3a and 3b support the hypothesis that surface frequency of the whole Arabic word is a predictor of the speed at which a speaker will recognize the word in an auditory lexical decision experiment, for two verbal paradigms: Binyan VIII and Binyan X. This is manifested by significantly faster RTs for words with higher surface frequencies, regardless of their root frequencies. These results contrast with the results found in the previous chapter, Chapter 3, which found that RTs to auditorily-presented words were significantly affected by both base frequency of the root morpheme and surface frequency of the whole verb. These results were found in two experiments: Experiment 2a and Experiment 2b. However, the verbs in Experiments 2a and 2b were from a productive verbal paradigm: Binyan I. Given that Binyan VIII and Binyan X are not considered productive, it was hypothesized that decomposability of verbs from these binyanim would not manifest in the same way as those from a productive binyan. Results from Experiments 3a and 3b support the hypothesis that productivity of the binyan affects decomposition during lexical access, and that a hybrid model such as those proposed by Balling and

Baayen (2008); Bertram et al. (2000a); Taft and Nguyen-Hoan (2010) which allow for both a morpheme-driven path to lexical access as well as a whole-word mechanism in which words are lexically-listed in their whole forms. This model is characterized by Taft and Nguyen-Hoan (2010) as an intermediate level in which word bases are linked to their derived forms. As evidenced by results from Spanish (López-Villaseñor, 2012), among others, the strategy used to recognize a word during lexical access depends on the productivity of the derivational affix it contains. For affixes which are productive, simple word surface frequency effects are observable, but also the speed at which a word is recognized depends on the frequency of its base morpheme. For affixes which are unproductive, the word will only be recognized whole-form and there is no look-up of a word's constituent morphemes, hence no base frequency effect.

These results make it clear that a model proposing complete decomposition of every multimorphemic word into their constituent morphemes (such as Taft (1979)) is not a nuanced enough mechanism to account for differences in productivity. Furthermore, recall that Boudelaa and Marslen-Wilson (2011) found that root productivity was the driving factor in decomposability resulting in priming as evidenced by facilitated RTs to a target word during a visual priming experiment as well as a cross-modal priming experiment. In these studies, it was found that prime-target pairs which shared an unproductive root did not result in significantly facilitated RTs, even when the root was used in a productive nominal binyan, suggesting that nominal binyan productivity is not relevant to the decomposability of an Arabic noun into its constituent morphemes during lexical access, and that root productivity is the driving factor of a noun's decomposability.

However, results of Experiments 3a and 3b show that in an auditory lexical decision experiment, verbs from unproductive binyanim are processed as a function of their surface frequencies. This is consistent with holistic whole-form access in which words are not fully decomposable into their constituent morphemes. If verbs of unproductive binyanim were fully decomposable throughout the whole lexical access process, we would expect that frequency of the root morphemes would have a significant effect on RTs during lexical access, however, the current study presents evidence that only surface frequency had a significant effect on RTs. The current results are not aligned with predictions of obligatory decomposition models (Meunier and Segui, 1999; Taft et al., 1986; Wurm, 2000) which propose a mechanism that decomposes a word into all of its constituent morphemes during lexical access, given that base frequency did not facilitate RTs, supporting the hypothesis that verbs with less productive verbal affixes were not recognized as a function of the frequency of their constituent morphemes.

The fact that binyan productivity does not trigger full decomposition for Arabic nouns during visual masked priming and cross-modal priming but that binyan productivity does drive decomposition in auditory lexical access for verbs suggests multiple possibilities for future research. First, it is possible that although binyan productivity does not play a role in decomposition at early stages of lexical access, it does determine decomposability in later stages of lexical access. This is consistent with the fact that priming effects are indicative of a measurement early in the lexical access process. Another possibility is that the productivity of nominal derivational affixes in Arabic has a different effect on the decomposability of a word than does the productivity of a verbal derivational affix.

Together, the results of the experiments presented in this chapter as well as in chapter 3 are consistent with the hypothesis that not all morphologically complex words are stored and accessed in the mental lexicon in the same way. For some Arabic verbs investigated here, the results suggest that word recognition is accomplished by breaking the word down into its constituent morphemes and accessed via these individual morphemes, as evidenced by the frequency effects of the base morpheme facilitating response speed. For other verbs, the surface frequency of the whole word facilitated lexical access resulting in faster responses. The verbs differed based on the productivity of their derivational affixes, supporting a model in which productivity determines decomposability during word recognition.

5. Conclusion

This dissertation proposed that the productivity of verbal affixes affects whether or not speakers are sensitive to morpheme frequency during the process of auditory word recognition. Arabic verbs are composed of a noncontiguous root and pattern (binyan) morphemes, the former of which is the base morpheme containing the meaning of the word and latter of which is a derivational affix which contributes grammatical information. These 10 binyanim differ in their attested productivity, which has implications for lexical processing. Specifically, it has been found for studies in Spanish (López-Villaseñor, 2012), English, Finnish, and Dutch (Bertram et al., 2000a,b), among others, that words with unproductive affixes are processed differently than words with productive affixes. The productivity of the derivational affix predicted whether or not the word would be decomposed and accessed by all of its constituent morphemes during lexical access or whether it would be stored and therefore accessed whole-form without further decomposition. This was evidenced by the effects of frequency of the base morpheme of the word in words with productive derivational affixes. For words with productive affixes, simple surface frequency effects of the whole word are observed, as well as the fact that words are recognized at a speed depending on the frequency of their base morphemes. For words with unproductive affixes, words were recognized holistically and were not recognized via decomposition of the word into its constituent elements as evidenced by the presence of word frequency effects and the absence of base frequency effects.

In the current study, it was therefore hypothesized that verbs in a more productive binyan would show a facilitatory effect of base morpheme frequency whereas verbs in a less productive binyan would not show an effect of base morpheme frequency. Specifically it was hypothesized that words in a productive binyan with frequent roots would be processed faster than words with infrequent roots, and that the surface frequency of the word itself would also have an effect. In contrast, for verbs in a less productive binyan, only the surface frequency of the word would affect processing speed. This hypothesis was supported by the results of four auditory lexical decision experiments. In the first two experiments, Experiment 2a and Experiment 2b, verbs in a productive binyan were found to have a significant effect of both base frequency and the surface frequency of the word itself. Experiments 3a and 3b explored the effects of the both base and surface frequency during auditory lexical access, but for verbs in less productive binyanim. For these verbs, speed of lexical access was found to be affected by only the surface frequency of the word itself as frequency of the base morpheme was not found to have an effect. Collectively, these results support the hypothesis that speakers are sensitive to affix productivity during the lexical access process. More productive affixes are more easily separated from base morphemes whereas less productive affixes are not, resulting in the a whole-word access during auditory processing. These results are consistent with those presented by López-Villaseñor (2012) for Spanish as well as research in English by Baayen et al. (2007) and Dutch and Finnish by Bertram et al. (2000b). Across these languages, it has also been demonstrated that affix productivity affects storage and access.

In this concluding chapter, a full summary of the results of the lexical decision experiments is presented. There is also a brief summary of the creation of the corpus used as a resource for the lexical statistics utilized, and results of the lexical familiarity experiment validating the corpus as a representative sample of Jordanian are reviewed. Following the summaries of the experimental results, implications of these results are discussed and situated in theories of lexical access. Finally, directions for future research are considered and the final conclusion of this dissertation is presented.

5.1. Summary of corpus creation and familiarity experiment

The experiments presented in this dissertation were facilitated by the creation of a corpus of Jordanian Arabic. This corpus served as the source of lexical statistics used to determine frequency and productivity of words and morphemes in Jordanian. The stimuli for the lexical decision experiments summarized below are drawn from the counts from this corpus.

Given that colloquial dialects of Arabic lack a written standard, creation of a colloquial corpus of a large enough size to provide robust lexicality measures and morpheme frequency information is a non-trivial task. To create a corpus of Jordanian Arabic, I began by consolidating existing resources written in Jordanian Arabic, including transcripts of conversational speech. Then, I used an automatic classifier to determine which portions of existing mixed Levantine corpora were likely to be Jordanian, and added them to existing resources. I then augmented the corpus by crowdsourding additional transcription of audio from broadcast speech, and mined colloquial-heavy web sources of writing, including Twitter and web forums, as detailed in chapter 2.

To confirm the validity of the corpus as a representative sample of Jordanian Arabic, I considered the attested correlation between corpus-based measures of frequency and subjective measures of frequency provided by participants in a word familiarity experiment. As shown by Balota et al. (2001); LaCross (2014); Francom et al. (2010), subjective frequency in the form of familiarity judgments to words mirror frequency measurements taken from a written corpus. Words which are judged as more familiar tend to have higher frequency counts whereas words which are judged as less familiar tend to have lower frequency counts. Experiment 1 was performed to collect subject familiarity judgments and compare their distribution with the frequencies measured in the completed corpus. Contributors of the crowdsourcing platform who were located in Jordan participated in Experiment 1 by listening to individual words of Jordanian and providing judgments as to how often they use the word, following methodology from Balota et al. (2001). Final frequency counts from the corpus were found to correlate with the subjective familiarity judgments provided from participants in Experiment 1, providing evidence that the corpus is representative of the language and valid for a resource of word and morpheme frequency measures for the lexical decision experiments summarized below.

Following completion of Experiment 1, the corpus was stemmed and roots were extracted to determine root frequency. Finally, verbal patterns were extracted and counts of verbs in each binyan were made. Following Baayen (1993); Bauer (2001), from these counts, the productivity of each binyan was determined. Productivity was measured by counting the number of hapaxes (words which appear only once) in the corpus that belong to that binyan and calculating their proportion in relation to the total number of hapaxes in the corpus.

5.2. Summary of lexical decision experiments

A total of four lexical decision experiments were performed and their results reported in this dissertation. These experiments were conducted with native speakers in Amman, Jordan and the stimuli for these experiments were drawn from the corpus described above.

5.2.1. Experiment 2a: Productive binyan with varied frequency of base morpheme

Experiment 2a was an auditory lexical decision experiment that aimed to uncover the role of frequency of the base morpheme in a complex word with a productive affix. In this experiment, participants provided lexical decision on words which differed based on their base frequency. Reaction Times (RTs) were collected to determine if participants responded significantly faster to words with more frequent roots than words with less frequent roots. For Experiment 2a, the verbs belonged to Binyan I of the form /CaCaC/. Surface frequency was controlled across items. It was hypothesized that these verbs are decomposable into their constituent morphemes during lexical access due to the fact that Binyan I is quite productive. Under this hypothesis, the base morpheme would be accessible during the lexical access process and would affect speed of recognition.

The results of Experiments 2a indicated that base frequency predicts RT and that the root morpheme affected how quickly speakers were able to recognize the word. This was indicated by the significant facilitatory effect of base frequency on RTs. Words with more frequent roots were responded to more quickly than words with less frequent roots.

These results supported the proposed hypothesis that the base morpheme was available to speakers during lexical access and therefore that higher base frequency facilitated speed of recognition. To determine if the surface frequency of the whole word also predicted speed of recognition, Experiment 2b was performed with a similar construction: surface frequency varied across items, and base frequency was controlled. As summarized below, evidence was found that both base and surface frequency are predictors of RT for Binyan I verbs.

5.2.2. Experiment 2b: Productive binyan with varied frequency of surface word

The results of Experiment 2a above demonstrated that base frequency facilitated RTs in a lexical decision task. For Experiment 2b, base frequency was controlled across items whereas the surface frequency of the whole word varied. This experiment also contained Binyan I verbs and RTs were analyzed to determine if surface frequency affected speed of recognition.

The effects of word frequency on RT in a lexical decision experiment have been well-attested (Howes, 1954; Newbigging, 1961; Savin, 1963; Whaley, 1978; Balota and Chumbley, 1984). More frequent words are responded to more quickly than less frequent words, and for multimorphemic words, when base frequency is held constant, words with high surface frequency are recognized more quickly than multimorphemic words with lower surface frequency (Taft, 1979; Burani et al., 1984; Colé et al., 1989). The results of Experiment 2b were found to be consistent with previous literature demonstrating that multimorphemic words are processed as a function of the frequency of the whole word.

Given that Experiments 2a and 2b contain verbs from a productive binyan, further experiments were performed to determine the role of base frequency in a multimorphemic word with a less productive binyan. These experiments are summarized below, and results showed that in contrast with Experiments 2a and 2b, multimorphemic words with a less productive binyan trigger whole-form storage and the constituent morphemes of the word do not play a role in lexical access.

5.2.3. Experiment 3a: Less productive binyan with varied base and surface frequency

Experiments 2a and 2b, as summarized above, presented evidence that words in a productive verbal paradigm, Binyan I, a processed by their frequency, as well as the frequency of their base morphemes. In contrast, Experiment 3a tested the hypothesis that verbs belonging to a less productive binyan: Binyan VIII of the form /iCtaCaC/. In this experiment, both surface and base frequency were manipulated for test items simultaneously. Given that this verbal affix is less productive, it was hypothesized that verbs containing this affix would be stored holistically and not be decomposed into their constituent morphemes during the lexical access process. Therefore, it was hypothesized that the frequency of the base morpheme for Binyan VIII words would not affect RTs, but that surface frequency of the whole word would facilitate RTs.

Results of Experiment 3a indicated that multimorphemic Arabic verbs are processed as a function of their surface frequency only, and not their base frequency.

This was evidenced by the facilitatory effect of surface frequency on RT, and no effect of base frequency on RT for Binyan VIII words. These results indicate that base morphemes are not accessible during the lexical access process for verbs with unproductive verbal affixes, and that these words are stored and accessed in their whole forms. Further evidence that full decomposability during lexical access is tied to affix productivity is presented below, in the summary of Experiment 3b.

5.2.4. Experiment 3b: Less productive binyan with varied base and surface frequency

In addition to the results presented above for Experiment 3a, further evidence that affix productivity determines decomposability of multimorphemic verbs is found in Experiment 3b, which sought to determine the role of base frequency and surface frequency for another less productive verbal affix, that associated with Binyan X /staCCaC/. Like Experiment 3a, Experiment 3b manipulated base frequency and surface frequency simultaneously. It was hypothesized that base frequency would not have an effect on RT and that verbs with high frequency roots would not be recognized significantly faster than words with low frequency roots. Furthermore, it was hypothesized that surface frequency would be a significant predictor of RT in this experiment. Specifically, it was predicted that words which were more frequent would be recognized faster than words which were less frequent.

Results of Experiment 3b supported this hypothesis, as evidenced by the facilitatory effect of surface frequency and the lack of effect of base frequency. These results indicate that the base morpheme of verbs with unproductive affixes are not accessible during auditory word recognition.

5.3. Implications

Collectively, the results of the experiments presented in this dissertation indicate that speakers of Arabic utilize multiple word recognition strategies when listening to multimorphemic words in speech. Speakers may access a word based on its base morpheme, or may access the word whole-form without further decomposition into the words constituent morphemes. As summarized above, the results of Experiments 2a and 2b indicated that the speed at which words with a more productive verbal affix are processed is affected by the frequency of the word itself as well as the frequency of its base morpheme. In contrast, the results of Experiments 3a and 3b showed that verbs with a relatively less productive affix are processed only as a function of the frequency of the word itself. The implications of the results of these experiments are that affix productivity drives the availability of a word's base morpheme during the lexical access process.

5.4. Productivity and morpheme frequency

The results of this dissertation provide additional evidence for the well-attested psycholinguistic saliency of the Arabic root morpheme, as previously demonstrated by Boudelaa and Marslen-Wilson (2004b, 2005, 2011) among others. However, in addition to the evidence presented in these previous studies, the current study indicates that the root morpheme is not accessible throughout the lexical access process for all Arabic verbs.

These findings provide evidence that the decomposability of Arabic verbs during lexical access is tied to the productivity of the morphological pattern. This contrasts with previous findings by Boudelaa and Marslen-Wilson (2011) concerning Arabic nouns. For Arabic nouns, it was found that the productivity of the root morpheme determined priming in a cross-modal and visual priming task. Specifically, if a prime-target pair shared a productive root, then response to the target was facilitated. In contrast, if a prime-target pair shared an unproductive root, then priming did not occur and the response was not facilitated. In this case, the productivity of the binyan was irrelevant and even a prime-target pair which shared a very productive binyann did not result in facilitated RTs if the root of the word was not productive. However, the results from this dissertation indicated that for Arabic verbs, binyan productivity does play a role in the accessibility of the root during lexical access.

These results are consistent with previous research showing that words are processed as a function of their base frequency. Evidence has been found demonstrating this in English (Baayen et al., 1997; Ford et al., 2010), Dutch (Baayen et al., 1997) and Italian (Burani et al., 1984). The evidence for English suggests that only words which were productively suffixed were affected by frequency of the base morpheme, and that words which lacked productive suffixes were not (Ford et al., 2010). It has also been demonstrated that the speed at which multimorphemic words are recognized is dependent on which derivational affix the word contained, in Dutch and English (Baayen et al., 1997).

It was hypothesized that the experimental results presented here would support a hybrid model of lexical access, such as those proposed by Baayen et al. (2007); Bertram et al. (2000b); Taft and Nguyen-Hoan (2010).
Within this framework, morphologically complex words are not recognized via a single method during lexical access. Instead, there are multiple paths to lexical access, including a morpheme-driven path in which a word is recognized by being decomposed into the morphemes it contains and a whole-word mechanism in which words are accessed in their whole forms holistically without further decomposition into the word's constituent morphemes. Results of the lexical decision experiments presented in this dissertation indicate support for these conclusions that holistic access is triggered by low derivational affix productivity.

These results do not indicate support for models proposing complete decomposition of all multimorphemic words (such as those proposed by Taft (1979, 2004);

Meunier and Segui (1999); Taft et al. (1986); Wurm (2000)). Under obligatory decomposition models, there is a role for all morphemes in lexical access, and no path for holistic storage. However, as presented above, the results presented in this dissertation show that for words with low-frequency affixes, a higher base frequency did not facilitate RTs. This indicates that the base morpheme is not accessible during lexical access as it has no effect on speed of recognition.

Furthermore, the results presented here are not consistent with previous work presented by Taft (2004) which showed evidence for an inhibitory effect for base frequency in words with low surface frequency. Taft (2004) presents this evidence in support of an obligatory decomposition model in which a word is decomposed into its constituent morphemes, given that words with high base frequency would require recombination of affix and base after decomposition during lexical access, which explains the effect of inhibition. This is presented as evidence against a hybrid model. However, results from the current study do not exhibit a reverse base frequency effect as words with higher base frequency show facilitated lexical access if they have a more productive affix, and for words with a less productive affix, there is a lack of base frequency effect.

The results of these experiments support a model such as that shown in Figure 5.1. A word with a productive affix would be recognized via decomposition, as shown in (a). In (a), both the whole word form and the intermediary base are available during access. The frequency of the base morpheme would thus have an effect on recognition speed, in addition to the processing effects from the frequency of the whole word. In contrast, in model (b), there is no access of the base morpheme. The word is accessed holistically.

In summary, it was hypothesized that the lexical decision experiments performed here would present evidence for both holistic word recognition and word recognition as a function of the base morpheme, in support of hybrid models of lexical access. The results of the lexical decision experiments did provide evidence for both whole-form

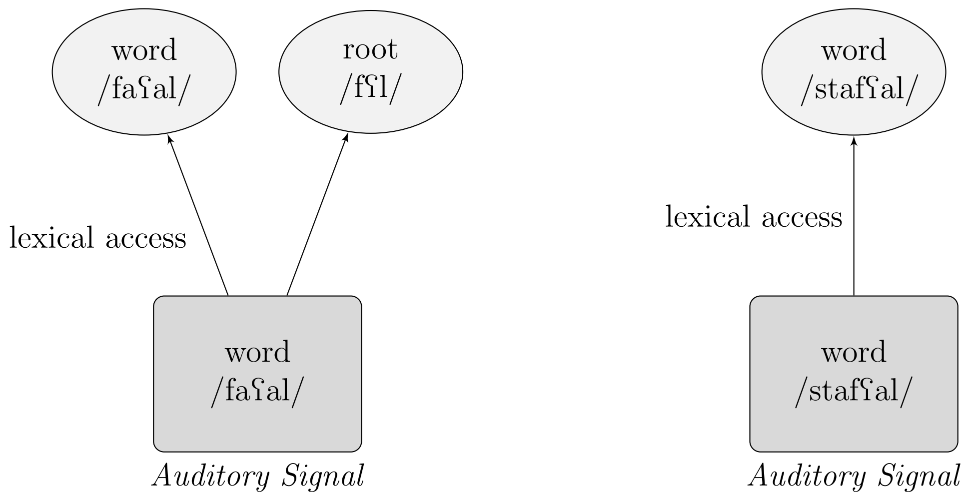

Figure a: Access with decomposition. Illustrates processing pathways for words with more productive affixes

Figure b: Holistic access. Illustrates processing pathway for words with less productive affixes

Figure 5.1: Hybrid model of auditory lexical access

and morpheme effects, and did support a hybrid model of auditory word recognition. As these models entail multiple potential pathways to recognize a word, evidence for different strategies of word recognition are appropriate for these models.

5.5. Directions of future research

The implications of this research raise potential questions for further research. Considering that Boudelaa and Marslen-Wilson (2011) found evidence that root productivity drives priming, whereas the current study shows that binyan productivity determines the availability of the root morpheme during lexical access, there are several possible implications in the differing results between these two studies. Although Boudelaa and Marslen-Wilson (2011) did not find evidence of binyan productivity driving decomposition, this may simply suggest the possibility of binyan productivity playing a role at the early stages of lexical access, but not determining decomposability during later stages of lexical access. Given that priming effects are indicative of early-stage effects during lexical access and that previous studies on binyan productivity including Boudelaa and Marslen-Wilson (2011) utilized priming as an experimental method, this may be an explanation for the lack of binyan productivity effect. It is also possible that nominal pattern effects may differ from verbal pattern effects. Together, these results suggest the need for further research, and further experimental evidence is necessary to gain a full picture of the roles of morphology in lexical access. For example, is the productivity of verbal patterns similarly irrelevant in priming when compared to the productivity of the root? Furthermore, what determines the recognition speed when speakers listen to nouns in a non-priming lexical decision task? Does nominal binyan productivity play a role in the speaker's ability to attend to the root in lexical access? It is important to consider too that Boudelaa and Marslen-Wilson (2011) define a morpheme's productivity as its family size, whereas the current study considers corpus-based measures of morpheme productivity. Future research will entail interpretation of the current results under the framework of morphological family size.

Another direction for future research concerns the existence of homonymy in the Arabic pattern morpheme. In English, Finnish and French, affix homonymy has been shown to trigger whole-form storage during lexical access (Bertram et al., 2000a, 1999). However, studies of pattern priming in Arabic do not show this effect: although /CiCa:Ca/ 'profession' and /CiCa:Ca/ 'deverbal noun' are homonymous morphemes, /tidʒa:ra/ 'trade' still primes /tˤiba:ʕa/ 'art of typography' (Boudelaa, 2014) . Their shared meaning of 'profession' still facilitates and the pair seemingly doesnt suffer any interference from the presence of homonyms of /CiCa:Ca/ in the language. However, homonyms do not prime each other, thus the word /ħika:ja/ 'story' inhibits /tˤiba:ʕa/ 'art of typography' (Boudelaa, 2014). Given these effects, further research is necessary to elucidate the nature of the pattern and separating out effects of productivity from effects of homonymy.

Furthermore, this dissertation has primarily focused on one kind of Arabic roots: triconsonantal strong roots. Strong roots are those which retain all three consonants in every conjugation. In contrast, weak roots are those with root consonants that surface as glides in some conjugations, and as long vowels in other. Although Schluter (2013) demonstrated a priming effect for both strong and weak roots in spoken Moroccan Arabic, further research is needed to determine if the recognition speed of weak roots is predicted only by the frequency of the word it contains, or if the frequency of the root predicts recognition speed as well.

5.6. Conclusion

This dissertation focused on the following question: how do Arabic speakers recognize multimorphemic words in speech? Specifically, are Arabic verbs recognized as a function of their frequency alone, or by the frequency of the base morphemes they contain? In this study, I have presented evidence that some verbs of Arabic are accessed holistically and some verbs are recognized by decomposition of a the word into its constituent morphemes. This dissertation has provided results supporting the role of affix productivity in lexical access. This dissertation also has provided data from Jordanian, a colloquial Levantine dialect which has been under-studied in psycholinguistics, despite having been well-studied phonologically and morphologically.

These results provided further elucidation to the nature of the Arabic lexicon and the factors influencing speed of lexical access for speakers of Arabic. Although numerous studies have demonstrated the decomposability of Arabic words during lexical access as evidenced partly by the psycholinguistic saliency of the root, results from this dissertation suggest that a clear role for binyan productivity in the process of lexical access is separate from the root.

Appendix A. Stimuli for Experiment 1

Word	Gloss	Word	Gloss
taʔkad	"he made sure"	nakab	"he subjected to distress"
istarsal	"he/it was free"	ħal:al	"he analyzed"
istaɣrab	"he/it found strange"	ixtalaf	"he/it differed"
ʔaras	"he was stinging"	fa:t	"he entered"
ixtaraʕ	"he/it suggested"	ʔadar	"he was able"
bid:o	"he wants"	istadʒmaʕ	"he gathered"
taʕlim	"he/it learned"	istaxdam	"he/it used"
iħtaram	"he respected"	wiʔaʕ	"he fell"
ibtaðal	"he violated good manners"	ixtalas	"he stole"
tafaradʒ	"he/it watched"	nahaʃ	"it tore"
istafsar	"he interrogated"	sˤadam	"it collided"
qaraʔ	"he read"	istabʕad	"he/it distanced"
xal:aa	"he/it left"	imtalak	"he obtained"
xal:asˤ	"he/it completed"	katab	"he/it wrote"
iʃtara	"he/it bought"	ɣadar	"he was treacherous"
istaɣal	"he/it made use of"	intaʔam	"he/it revenged"
iħmar	"he/it reddened"	irtakaz	"it was based on"
iʔtˤar	"he/it pulled"	ʃa:war	"he consulted"
iʃtahar	"he/it became famous"	tuwaʔfa	"he/it stopped"
fataʃ	"he/it searched"	iɣta:b	"he backbit"
laʕab	"he/it played"	xaðal	"he disappointed"
iħtaʔar	"he humiliated"	naʕas	"he became sleepy"
iftaxar	"he/it thought"	ʃa:f	"he saw"
xalaʕ	"he/it finished"	tbar:aʕ	"he donated"
ʃaras	"he became aggressive"	haðˤam	"he/it digested"
ixtaraʕ	"he invented"	xabar	"he/it told"
imsik	"he/it grabbed"	ihtaʔar	"he humiliated"
na:m	"he slept"	baðal	"he/it extended"
balaʃ	"he started"	faxam	"it carbonized"
iʕtaraf	"he admitted"	istanbatˤ	"he/it derived"

Appendix B. Language History Survey

1. Are you a native speaker of Arabic?
2. What is your native dialect?
   (a) Jordanian
   (b) Palestinian
   (c) Other:
3. What city do you live in?
4. What city were you born in?
5. What city did you grow up in?
6. How many months of the year do you spend in Jordan?
7. Have you ever lived outside Jordan? If so, list the countries you have lived in:
8. What was your primary language of schooling for your primary school?
   (a) Arabic
   (b) English
   (c) French
   (d) Other:
9. What was your primary language of schooling for your middle school?
   (a) Arabic
   (b) English

(c) French

(d) Other:

10. What was your primary language of schooling for your high school?

    (a) Arabic

    (b) English

    (c) French

    (d) Other:

11. Did you attend University?

    If yes, how many years?

    Which university did you attend?

    What was the primary language of schooling at university?

12. In what dialect do you speak to the following people:

    (a) Mother

    (b) Father

    (c) Grandparents

    (d) Siblings

    (e) Uncles + aunts

    (f) Children

    (g) Friends

    (h) Coworkers

    (i) Customers/clients

13. In which language/dialect are the newspapers and books you typically read written? Indicate a percentage:

(a) MSA: %

(b) English: %

(c) Other (please specify:): %

(d) Other (please specify:): %

14. In which language/dialect are the TV programs you typically watch spoken? Indicate a percentage:

    (a) MSA: %

    (b) Jordanian: %

    (c) Palestinian: %

    (d) Syrian: %

    (e) Egyptian: %

    (f) English: %

    (g) Other (please specify:): %

    (h) Other (please specify:): %

15. In which language/dialect are the radio programs you typically listen to spoken? Indicate a percentage:

    (a) MSA: %

    (b) Jordanian: %

    (c) Palestinian: %

    (d) Syrian: %

    (e) Egyptian: %

    (f) English: %

    (g) Other (please specify:): %

(h) Other (please specify:): %

16. How often do you chat/text in colloquial dialect?

    (a) Multiple times a day

    (b) One a week

    (c) A few times a year

    (d) Never

Appendix C. Stimuli for Experiment 2a

Word	Gloss	Word	Gloss
ħarasˤ	"he/it was keen"	nafat	"he/it coughed"
taʔil	"he/it became heavy"	dʒaraʃ	"he/it crushed"
miraħ	"he/it had fun"	ta:ʔ	"he/it craved"
tˤa:r	"he/it flew"	tˤifar	"he/it went broke"
ʕazam	"he/it was intent"	dʒahar	"he/it was known"
ħaraʃ	"he/it incited"	habaʃ	"he/it gathered"
balatˤ	"he/it tiled"	xaratˤ	"he/it lied"
ʃabak	"he/it linked"	ʔatˤam	"he/it cut"
baʕat	"he/it sent"	sˤaʕa'	"he/it was shocked"
ɣadar	"he/it betrayed"	fasaʔ	"he/it acted immorally"
xabat	"he/it became malicious"	taledʒ	"he/it froze"
katam	"he/it concealed"	sˤafad	"he/it enchained"
ʕaraʔ	"he/it sweat"	hamaz	"he/it back bited"
sˤadam	"he/it bumped"	waħil	"he/it became stuck (in mud)"
ħalal	"he/it analyzed"	ʔadˤim	"he/it bit"
lamaħ	"he/it glanced"	taʕas	"he/it became miserable"
ʕasˤab	"he/it became angry"	sˤafin	"he/it stared"
ʃaʕal	"he/it burned"	madˤaɣ	"he/it chewed"
fatan	"he/it seduced"	ʃarax	"he/it cracked"
nisum	"he/it began"	nabaz	"he/it exited"
risax	"he/it was deep"	tabatˤ	"he/it intimidated"
ratim	"he/it spoke cryptically"	xadˤam	"he/it cut"
ħadˤar	"he/it banned"	sijaʕ	"he/it murmured"
nabatˤ	"he/it showed"	sifaʕ	"he/it slapped"
karaʕ	"he/it drank"	sˤaxab	"he/it clamored"
xasaf	"he/it collapsed"	zaʔam	"he/it swallowed"
ɣalaf	"he/it encased"	ʔazam	"he/it incapacitated"
saħil	"he/it slipped"	dʒaħadh	"he/it protruded"
nabaħ	"he/it barked"	ʕaʃam	"he/it was parched"
xaras	"he/it shut up"	samatˤ	"he/it quieted"

Appendix D. Stimuli for Experiment 2b

Word	Gloss
rasab	“he/it failed”
ʔarasˤ	“he/it stung”
tˤarib	“he/it sung”
raʕaʃ	“he/it bounced”
sumim	“he/it poisoned”
fataʃ	“he/it searched”
nasim	“he/it began”
kabas	“he/it pressed”
daliʔ	“he/it poured”
xalaʕ	“he/it removed”
nakab	“he/it distressed”
sixar	“he/it teased”
satˤal	“he/it startled”
nadʒid	“he/it upholstered”
batˤar	“he/it disregarded”
ʕasar	“he/it roughened”
zakar	“he/it reminded”
ħasar	“he/it became brokenhearted”
madaħ	“he/it complimented”
nafax	“he/it swole”
daaʔ	“he/it tasted”
ridʒaʕ	“he/it preferred”
fiʔaʕ	“he/it burst”
duhiʃ	“he/it was surprised”
thaʕar	“he/it was terrified”
tˤafaʃ	“he/it left”
ʃatim	“he/it insulted”
taʔib	“he/it punctured”
dʒinaħ	“he/it departed”
ħasˤad	“he/it reaped”

Word	Gloss
nafal	“he/it made a mess”
zamar	“he/it honked”
ʕatim	“he/it covered”
rihan	“he/it paid security”
ħazar	“he/it gussed”
tˤarid	“he/it expelled”
habat	“he/it fell”
raʔid	“he/it slept”
ħabas	“he/it imprisoned”
xaðal	“he/it disappointed”
faħam	“he/it carbonized”
ħatˤab	“he/it gathered wood”
hadˤam	“he/it digested”
ʃaniʕ	“he/it strangled”
rafaʃ	“he/it knocked”
xasˤal	“he/it provided”
shatˤib	“he/it deleted”
maʃat	“he/it combed”
ʕabat	“he/it abused”
ʔatˤab	“he/it sewed”
ramash	“he/it flickered”
faxam	“he/it regarded”
nahab	“he/it stole”
baðal	“he/it exerted”
sakab	“he/it poured”
naʕas	“he/it became sleepy”
kasal	“he/it was sluggish”
ramaħ	“he/it pierced”
salax	“he/it peeled skin”
sˤahar	“he/it liquefied”

Appendix E. Stimuli for Experiment 3a

Word	Gloss
iltaʔam	"he/it devoured"
ibtaðal	"he/it violated manners"
iʔtatˤaf	"he/it cut and gathered"
irtakaz	"he/it was based on"
iʔtasam	"he/it shared"
iltamas	"he/it commanded"
iʃtabah	"he/it was similar"
iʃtabak	"he/it was complicated"
iktasaħ	"he/it invaded"
iʔtaraf	"he/it committed"
imtahan	"he/it was professional"
iħtaʔar	"he/it degraded"
iħtabas	"he/it was imprisoned"
intaʃal	"he/it snatched"
ixtalas	"he/it stole"
iʕtaʔal	"he/it arrested"
iftataħ	"he/it started"
iftakar	"he/it remembered"
iʔtatˤaʕ	"he/it cut"
ixtabar	"he/it tested"
iʔtasˤad	"he/it economized"
imtalaʔ	"he/it became full"
iħtasab	"he/it considered"
iɣtarab	"he/it went abroad"
iħtaraf	"he/it was professional"
iftaraʔ	"he/it became separated"
istalam	"he/it received"
iħtafadˤ	"he/it saved or kept"
iɣta:b	"he/it backbit"
iftaʕal	"he/it fabricated"

Word	Gloss
intaʔam	"he/it took revenge"
irtaʕaʃ	"he/it tingled"
intasˤaf	"he appealed for justice"
intaħar	"he/it committed suicide"
iʕtasˤam	"he/it adhered'
intazaʕ	"he/it seized"
irtadʒaf	"he/it dithered"
idʒtahad	"he/it worked hard"
ixtaraʔ	"he/it permeated"
iltaʔatˤ	"he/it caught"
ixtalatˤ	"he/it mixed"
iʔtanaʕ	"he/it became convinced"
ixtaraʕ	"he/it invented"
iʔtaraħ	"he/it suggested"
iftaxar	"he/it became proud"
irtafaʕ	"he/it ascended"
ibtaʕad	"he/it fled"
iktamal	"he/it accomplished"
intasab	"he/it was connected with"
iltazam	"he/it committed"
iħtaram	"he/it respected"
imtalak	"he/it obtained"
intaðˤar	"he/it waited"
iʃtahar	"he/it became well known"
iʔtarab	"he/it got near"
idʒtamaʕ	"he/it accumulated"
ixtalaf	"he/it argued"
iʃtaɣal	"he/it worked"
iʕtaraf	"he/it admitted"
iʕtabar	"he/it learned from"

Appendix F. Stimuli for Experiment 3b

Word	Gloss
staħwað	“he/it was on a mind”
stanbatˤ	“he/it drew out”
staħam	“he/it took a shower”
stafaz	“he/it annoyed”
stayʔaðˤ	“he/it woke up”
standʒad	“he/it appealed for help”
stafarɣ	“he/it vomited”
stawħaʃ	“he/it was alone”
staʕtˤaf	“he/it asked for compassion”
stabraʔ	“he/it embroidered”
staʕa:r	“he/it borrowed”
staħlaf	“he/it required an oath”
starzaʔ	“he/it sought gain”
starˤa:ʕ	“he/it could”
staʕdʒal	“he/it quickened”
staʕad:	“he/it prepared”
sta:hal	“he/it deserved”
stawtˤan	“he/it took up residence”
stasˤʕab	“he/it found difficult”
stafham	“he/it inquired”
stakmal	“he/it completed”
stafa:ʔ	“he/it woke up”
staʕad	“he/it regarded as unlikely”
stalzam	“he/it did necessary”
staħdat	“he/it created”
staxlasˤ	“he/it drew conclusion”
stama:l	“he/it persuaded”
staʔbal	“he/it accepted”
stakthar	“he/it deemed excessive”
stadʒmaʕ	“he/it gathered”

Word	Gloss
staxaf:	“he/it underestimated”
stawʕab	“he/it understood”
staʔdʒar	“he/it rented”
statmar	“he/it invested”
staɣal:	“he/it benefited”
stafsar	“he/it asked explanation”
staɣfar	“he/it apologized”
staħaʔ	“he/it deserved”
stahbal	“he/it foolish”
stamtaʕ	“he/it enjoyed”
stawdaʕ	“he/it put up a deposit”
stahdaf	“he/it aimed for”
stabdal	“he/it was replaced”
stafa:d	“he/it benefited”
staʔnas	“he/it had fun”
staðˤa:f	“he/it received a guest”
staʕraðˤ	“he/it inspected”
staɣrab	“he/it found strange”
staxdam	“he/it used”
stawʔaf	“he/it asked to stop”
staʃhad	“he/it became a martyr”
staxlaf	“he/it appointed”
staħmal	“he/it tolerated”
stadʒa:b	“he/it responded to”
staħsan	“he/it considered good”
starsal	“he/it went too far”
staʔa:m	“he/it behaved well”
staslam	“he/it surrendered”
stada:r	“he/it turned around”
stamar:	“he/it continued”

References

Abu-Haidar, F. (1989). Are Iraqi women more prestige conscious than men? Sex differentiation in Baghdadi Arabic. *Language in Society*, 18(04):471–481.

Al-Sabbagh, R. and Girju, R. (2012). YADAC: Yet another Dialectal Arabic Corpus. In *LREC*, pages 2882–2889.

Al-Wer, E. (2003). New dialect formation: the focusing of–kum in Amman. *Social Dialectology: In honour of Peter Trudgill. Amsterdam: Benjamins*, pages 59–67.

Al-Wer, E. (2007). The formation of the dialect of Amman. *Arabic in the city: Issues in dialect contact and language variation*, 5:55.

Al-Wer, E. and Herin, B. (2011). The lifecycle of Qaf in Jordan. *Langage et société*, (4):59–76.

Alegre, M. and Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of memory and language*, 40(1):41–61.

Almeman, K. and Lee, M. (2013). Automatic building of Arabic Multi Dialect text corpora by bootstrapping dialect words. In *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on*, pages 1–6. IEEE.

Almeman, K., Lee, M., and Almiman, A. A. (2013). Multi dialect Arabic speech parallel corpora. In *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on*, pages 1–6. IEEE.

Appen (2007). *Levantine Arabic Conversational Telephone Speech LDC2007T01*. Linguistics Data Consortium.

Audhkhasi, K., Georgiou, P., and Narayanan, S. S. (2011). Accurate transcription of broadcast news speech using multiple noisy transcribers and unsupervised reliability metrics. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 4980–4983. IEEE.

Azban, A. K. (2011). *Diwan Baladna: The unprecedented spoken Arabic dictionary.*

Baayen, H. (1993). On frequency, transparency and productivity. In *Yearbook of Morphology 1992*, pages 181–208. Springer.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):390–412.

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

Badawi, A.-S. M. (1973). *Mustawayat al-arabiyya al-muasira fi Misr*. Dar al-maarif.

Bakr, H. A., Shaalan, K., and Ziedan, I. (2008). A hybrid approach for converting written Egyptian colloquial dialect into diacritized Arabic. In *The 6th international conference on informatics and systems, infos2008. cairo university*.

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

Balota, D. A. and Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception and performance*, 10(3):340.

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

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Bassiouney, R. (2009). *Arabic sociolinguistics*. Edinburgh University Press.

Bates, D. (2005). Fitting linear mixed models in r. r news 5: 27–30. *See http://cran. rproj ect. org/doc/Rnews/RnewsfZO05-1. pdf (accessed 17 May 2011)*.

Bauer, L. (2001). *Morphological productivity*, volume 95. Cambridge University Press.

Berent, I., Vaknin, V., and Marcus, G. F. (2007). Roots, stems, and the universality of lexical representations: Evidence from Hebrew. *Cognition*, 104(2):254–286.

Bertram, R., Baayen, R. H., and Schreuder, R. (2000a). Effects of family size for complex words. *Journal of Memory and Language*, 42(3):390–405.

Bertram, R., Laine, M., and Karvinen, K. (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of psycholinguistic research*, 28(3):213–226.

Bertram, R., Schreuder, R., and Baayen, R. H. (2000b). 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.

Bouamor, H., Habash, N., and Oflazer, K. (2014). A Multidialectal Parallel Corpus of Arabic. In *LREC*, pages 1240–1245.

Boudelaa, S. (2014). Is the Arabic mental lexicon morpheme-based or stem-based? Implications for spoken and written word recognition. In *Handbook of Arabic literacy*, pages 31–54. Springer.

Boudelaa, S. and Marslen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing: Evidence from Modern Standard Arabic. In *ISCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes*.

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

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

Boudelaa, S. and 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. and Marslen-Wilson, W. D. (2010). Aralex: A lexical database for modern standard Arabic. *Behavior Research Methods*, 42(2):481–487.

Boudelaa, S. and Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes*, 26(4-6):624–652.

Boudelaa, S. and 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.

Bradley, D. (1980). Lexical representation of derivational relation. *Juncture*, pages 37–55.

Brysbaert, M. and New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, 41(4):977–990.

Buckwalter, T. (2003). Buckwalter Arabic Corpus.

Buckwalter, T. (2004). Buckwalter Arabic morphological analyzer (BAMA) version 2.0. linguistic data consortium (LDC) catalogue number LDC2004L02. Technical report, ISBN1-58563-324-0.

Buckwalter, T. (2007). Issues in Arabic morphological analysis. In *Arabic computational morphology*, pages 23–41. Springer.

Burani, C. and Caramazza, A. (1987). Representation and processing of derived words. *Language and cognitive processes*, 2(3-4):217–227.

Burani, C., Salmaso, D., and Caramazza, A. (1984). Morphological structure and lexical access. *Visible Language*, 18(4):342–352.

Canavan, A. and Zipperlen, G. (1996). CALLFRIEND Egyptian Arabic. *LDC, Philadelphia*.

Canavan, A., Zipperlen, G., and Graff, D. (1997). CALLHOME Egyptian Arabic Speech. *Philadelphia, PA: Linguistic Data Consortium*.

Cantineau, J. (1950). *La notion de" schème" et son altération dans diverses langues sémitiques*. Maisonneuve.

Cohen, M. (1951). Langues chamito-sémitiques et linguistique historique.

Colé, P., Beauvillain, C., and 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.

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

Cotterell, R. and Callison-Burch, C. (2014). A multi-dialect, multi-genre corpus of informal written Arabic. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*.

crowdflower.com (2015). Crowdflower.

Daher, J. (1998). Gender in linguistic variation: The variable (q) in Damascus Arabic. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, pages 183–208.

Darwish, K., Magdy, W., and Mourad, A. (2012). Language processing for Arabic microblog retrieval. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 2427–2430. ACM.

Davis, S. (2011). Geminates. *The Blackwell companion to phonology*, 2:837–859.

Davis, S. and Zawaydeh, B. A. (2001). Arabic hypocoristics and the status of the consonantal root. *Linguistic Inquiry*, 32(3):512–520.

De Jong, N. H., Schreuder, R., and Harald Baayen, R. (2000). The morphological family size effect and morphology. *Language and cognitive processes*, 15(4-5):329–365.

del Prado Martín, F. M., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., and 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.

Deutsch, A., Frost, R., and 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.

Diab, M. and Habash, N. (2007). Arabic dialect processing tutorial. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Tutorial Abstracts*, pages 5–6. Association for Computational Linguistics.

Diab, M. and Habash, N. (2014). Natural Language Processing of Arabic and its Dialects.

Domínguez, A., Cuetos, F., and Segui, J. (2000). Morphological processing in word recognition: A review with particular reference to Spanish data. *Psicológica*, 21(2):375–401.

Duh, K. and Kirchhoff, K. (2005). POS tagging of dialectal Arabic: a minimally supervised approach. In *Proceedings of the acl workshop on computational approaches to semitic languages*, pages 55–62. Association for Computational Linguistics.

Eisenstein, J. (2013). What to do about bad language on the internet. In *HLT-NAACL*, pages 359–369.

El-Defrawy, M., El-Sonbaty, Y., and Belal, N. A. (2015). Enhancing Root Extractors Using Light Stemmers.

Elfardy, H. and Diab, M. T. (2013). Sentence level dialect identification in arabic. In *ACL (2)*, pages 456–461.

Elihay, J. (2004). *The Olive Tree Dictionary: A Transliterated Dictionary Of Conversational Arabic*. Kidron Pub.

Elmahdy, M., Hasegawa-Johnson, M., and Mustafawi, E. (2014). Development of a TV Broadcast Speech Recognition System for Qatari Arabic. In Chair), N. C. C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

ELRA (2014). *A-SpeechDB v. 1.0.*

Emmorey, K. D. (1989). Auditory morphological priming in the lexicon. *Language and Cognitive Processes*, 4(2):73–92.

Eskander, R., Habash, N., Rambow, O., and Tomeh, N. (2013). Processing Spontaneous Orthography. In *HLT-NAACL*, pages 585–595.

Ferguson, C. A. (1959). The Arabic Koine. *Language*, pages 616–630.

Ford, M., Davis, M., and Marslen-Wilson, W. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language*, 63(1):117–130.

Forster, K. I. and Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1):116–124.

Francom, J., LaCross, A., and Ussishkin, A. (2010). How Specialized are Specialized Corpora? Behavioral Evaluation of Corpus Representativeness for Maltese. In *LREC*.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., and Marslen-Wilson, W. (2000). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory & Cognition*, 28(8):1277–1288.

Frost, R., Forster, K. I., and 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.

Gaskell, M. G. and Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive psychology*, 45(2):220–266.

Gelas, H., Abate, S. T., Besacier, L., and Pellegrino, F. (2011). Quality assessment of crowdsourcing transcriptions for African languages. In *Interspeech*.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of experimental psychology: General*, 113(2):256.

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes*, 11(6):559–568.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., and Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6):1211.

Gonnerman, L. M., Seidenberg, M. S., and 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.

Gordon, P. and Alegre, M. (1999). Is there a dual system for regular inflections? *Brain and Language*, 68(1):212–217.

Grainger, J., Muneaux, M., Farioli, F., and Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology Section A*, 58(6):981–998.

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

Habash, N., Diab, M. T., and Rambow, O. (2012a). Conventional Orthography for Dialectal Arabic. In *LREC*.

Habash, N., Eskander, R., and Hawwari, A. (2012b). A morphological analyzer for Egyptian Arabic. In *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology*, pages 1–9. Association for Computational Linguistics.

Habash, N., Rambow, O., and Roth, R. (2009). MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In *Proceedings of the 2nd international conference on Arabic language resources and tools (MEDAR), Cairo, Egypt*, pages 102–109.

Habash, N., Roth, R., Rambow, O., Eskander, R., and Tomeh, N. (2013). Morphological Analysis and Disambiguation for Dialectal Arabic. In *HLT-NAACL*, pages 426–432.

Hämäläinen, A., Pinto Moreira, F., Avelar, J., Braga, D., and Sales Dias, M. (2013). Transcribing and Annotating Speech Corpora for Speech Recognition: A Three-Step Crowdsourcing Approach with Quality Control. In *First AAAI Conference on Human Computation and Crowdsourcing*.

Hary, B. (1996). The importance of the language continuum in Arabic multiglossia. *Understanding Arabic*, pages 69–90.

Hetzron, R. and Bender, M. L. (1976). The Ethio-Semitic Languages. *Languages in Ethiopia, ed. ML Bender et al*, pages 23–33.

Hilaal, Y. (1990). Deriving from roots and word patterns. *Linguistica Communicatio*, 1:77–80.

Holes, C. (1996). The Arabic Dialects of South Eastern Arabia in a socio-historical perspective. *Zeitschrift für arabische Linguistik*, (31):34–56.

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology*, 48(2):106.

Huck, J., Whyatt, D., and Coulton, P. (2012). Challenges in geocoding socially-generated data.

Ibrahim, M. H. (1983). Linguistic distance and literacy in Arabic. *Journal of Pragmatics*, 7(5):507–515.

Iskra, D. J., Siemund, R., Borno, J., Moreno, A., Emam, O., Choukri, K., Gedge, O., Tropf, H. S., Nogueiras, A., Zitouni, I., et al. (2004). OrienTel-Telephony Databases Across Northern Africa and the Middle East. In *LREC*. Citeseer.

Jastrow, O. (1980). Die dialekte der arabischen Halbinsel. *Fischer and Jastrow*.

Joanisse, M. F. and Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences*, 96(13):7592–7597.

Kaye, A. S. (1994). Formal vs. Informal in Arabic: Diglossia, Triglossia, Tetraglossia, etc., PolyglossiaMultiglossia Viewed as a Continuum. *Zeitschrift für arabische Linguistik*, (27):47–66.

Kaye, A. S. (2007). *Morphologies of Asia and Africa: Volume 1*. Eisenbrauns.

Kempley, S. T. and Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, 73(4):441–454.

Khoja, S. and Garside, R. (1999). Stemming Arabic Text. *Lancaster, UK, Computing Department, Lancaster University*.

Kouider, S. and Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, 16(8):617–625.

LaCross, A. (2014). Khalkha Mongolian speakers' vowel bias: L1 influences on the acquisition of non-adjacent vocalic dependencies. *Language, Cognition and Neuroscience*, pages 1–15.

Laks, L. (2009). The formation of Arabic passive verbs: lexical or syntactic. In *Proceedings of IATL*, volume 25.

Lee, C. and Glass, J. (2011). A Transcription Task for Crowdsourcing with Automatic Quality Control. In *Interspeech*, pages 3041–3044.

Lewis, G., Solomyak, O., and Marantz, A. (2011). The neural basis of obligatory decomposition of suffixed words. *Brain and language*, 118(3):118–127.

Ljubešic, N., Fišer, D., and Erjavec, T. (2014). Tweet-CaT: a Tool for Building Twitter Corpora of Smaller Languages. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC14), Reykjavik, Iceland. European Language Resources Association (ELRA).*

López-Villaseñor, M. L. (2012). The effects of base frequency and affix productivity in Spanish. *The Spanish journal of psychology*, 15(02):505–512.

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

Luce, R. D. (1986). *Response times.* Number 8. Oxford University Press.

Lukatela, G. and Turvey, M. T. (1994). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General*, 123(4):331.

Maamouri, M., Buckwalter, T., Graff, D., and Jin, H. (2007). *Fisher Levantine Arabic Conversational Telephone Speech LDC2007S02.* Linguistics Data Consortium.

Maamouri, M., Graff, D., Bouziri, B., Krouna, S., Bies, A., and Kulick, S. (2010). Standard Arabic morphological analyzer (SAMA) version 3.1. *Linguistic Data Consortium, Catalog No.: LDC2010L01.*

Makhoul, J., Zawaydeh, B., Choi, F., and Stallard, D. (2005). BBN/AUB DARPA Babylon Levantine Arabic speech and transcripts. *Linguistic Data Consortium (LDC), LDC Catalog No.: LDC2005S08.*

Malmasi, S., Refaee, E., and Dras, M. (2015). Arabic dialect identification using a parallel multidialectal corpus. In *Proceedings of the 14th Conference of the Pacific Association for Computational Linguistics (PACLING 2015), Bali, Indonesia*, pages 209–217.

Mansour, S., Al-Onaizan, Y., Blackwood, G., and Tillmann, C. (2014). Automatic Dialect Classification for Statistical Machine Translation. In *Proc. of the Conf. of the Assoc. for Machine Translation in the Americas (AMTA), Vancouver, BC, Canada*.

Marge, M., Banerjee, S., and Rudnicky, A. I. (2010a). Using the Amazon Mechanical Turk for transcription of spoken language. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5270–5273. IEEE.

Marge, M., Banerjee, S., and Rudnicky, A. I. (2010b). Using the Amazon Mechanical Turk for transcription of spoken language. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5270–5273. IEEE.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access.

Marslen-Wilson, W., Moss, H. E., and van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22(6):1376.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., and Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological review*, 101(1):3.

Marslen-Wilson, W. D., Tyler, L. K., et al. (1997). Dissociating types of mental computation. *Nature*, 387(6633):592–593.

Marslen-Wilson, W. D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, 10(1):29–63.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic inquiry*, 12(3):373–418.

McClelland, J. L. and Elman, J. L. (1986a). Interactive processes in speech perception: The TRACE model. *Parallel distributed processing*, 2(58):121.

McClelland, J. L. and Elman, J. L. (1986b). The TRACE model of speech perception. *Cognitive psychology*, 18(1):1–86.

McNeil, K. and Faiza, M. (2011). Tunisian Arabic Corpus: Creating a written corpus of an unwritten language. Presented at Workshop for Arabic Corpus Linguistics (WACL).

Meignier, S. and Merlin, T. (2010). LIUM SpkDiarization: an open source toolkit for diarization. In *CMU SPUD Workshop*.

Meunier, F. and Segui, J. (1999). Morphological priming effect: The role of surface frequency. *Brain and Language*, 68(1):54–60.

Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, 90(2):227.

Miller, C. (2007). Arabic urban vernaculars. *Arabic in the city: issues in dialect contact and language variation*, 5:1.

Mitchell, T. F. (1978). Educated spoken Arabic in Egypt and the Levant, with special reference to participle and tense. *Journal of Linguistics*, 14(02):227–258.

Mitchell, T. F. (1986). What is educated spoken Arabic? *International Journal of the sociology of language*, 61(1):7–32.

Mubarak, H. and Darwish, K. (2014). Using Twitter to collect a multi-dialectal corpus of Arabic. *ANLP 2014*, page 1.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., and Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, pages 262–282.

New, B., Brysbaert, M., Segui, J., Ferrand, L., and Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, 51(4):568–585.

Newbigging, P. L. (1961). The perceptual redintegration of frequent and infrequent words. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 15(3):123.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3):189–234.

Norris, D. and McQueen, J. M. (2008). Shortlist b: a bayesian model of continuous speech recognition. *Psychological review*, 115(2):357.

Novotney, S. and Callison-Burch, C. (2010). Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-expert Transcription. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 207–215, Stroudsburg, PA, USA. Association for Computational Linguistics.

Palfreyman, D. and Khalil, M. a. (2003). A Funky Language for Teenzz to Use: Representing Gulf Arabic in Instant Messaging. *Journal of Computer-Mediated Communication*, 9(1):0–0.

Parker, R., Graff, D., Chen, K., Kong, J., and Maeda, K. (2011). Arabic Gigaword fifth edition LDC2011T11. *Philadelphia: Linguistic Data Consortium*.

Pasha, A., Al-Badrashiny, M., Diab, M., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., and Roth, R. M. (2014). Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. In *Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland*.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2016). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-124.

Plaut, D. C. and 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.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, 103(1):56.

Prunet, J.-F., Béland, R., and Idrissi, A. (2000). The mental representation of Semitic words. *Linguistic inquiry*, 31(4):609–648.

R-Core-Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.

Radeau, M. (1983). Semantic priming between spoken words in adults and children. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 37(4):547.

Radeau, M., Morais, J., and Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition*, 17(5):525–535.

Radeau, M., Morais, J., and Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1297.

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-'Arabiyya*, pages 53–75.

Refaee, E. and Rieser, V. (2014). Subjectivity and sentiment analysis of Arabic twitter feeds with limited resources. In *Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools Workshop Programme*, page 16.

Richardson, L. (2008). Beautiful soup-html. *XML parser for Python*.

Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., and Mars, F. (1997). Morphological priming, fragment completion, and connectionist networks. *Journal of Memory and Language*, 36(3):382–405.

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

Savin, H. B. (1963). Word-Frequency Effect and Errors in the Perception of Speech. *The Journal of the Acoustical Society of America*, 35(2):200–206.

Schluter, K. T. (2013). Hearing words without structure: Subliminal speech priming and the organization of the Moroccan Arabic lexicon.

Schreuder, R. and Baayen, R. H. (1997). How complex simplex words can be. *Journal of memory and language*, 37(1):118–139.

Schriefers, H., Friederici, A., and Graetz, P. (1992). Inflectional and derivational morphology in the mental lexicon: Symmetries and asymmetries in repetition priming. *The Quarterly Journal of Experimental Psychology*, 44(2):373–390.

Seidenberg, M. S. and Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in cognitive sciences*, 4(9):353–361.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, 96(4):523.

Slowiaczek, L. M. and Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6):1239.

Slowiaczek, L. M. and Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, 14(3):230–237.

Smolka, E., Komlosi, S., and Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, 24(3):337–375.

Spieler, D. H. and Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, pages 411–416.

Stahlberg, F. and Vogel, S. (2015). The QCRI Recognition System for Handwritten Arabic. In *ICIAP*.

Stolcke, A. et al. (2002). SRILM - An extensible language modeling toolkit. In *INTERSPEECH*, volume 2002, page 2002.

Suleiman, S. M. (1985). *Jordanian Arabic between diglossia and bilingualism: Linguistic analysis*. John Benjamins Publishing.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, 7(4):263–272.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, 57(4):745–765.

Taft, M., Hambly, G., and Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *The Quarterly Journal of Experimental Psychology*, 38(3):351–365.

Taft, M. and Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes*, 25(2):277–296.

Tiedemann, F. E. (2015). *The 101 Most Used Verbs in Spoken Arabic: Jordan & Palestine*.

Tillmann, C., Al-Onaizan, Y., and Mansour, S. (2014). Improved sentence-level Arabic dialect classification. In *Proceedings of the VarDial Workshop*, pages 110–119.

Twist, A. E. (2006). A psycholinguistic investigation of the verbal morphology of Maltese.

Tyler, L. K., Anokhina, R., Longworth, C., Randall, B., Marslen-Wilson, W. D., et al. (2002). Dissociations in processing past tense morphology: Neuropathology and behavioral studies. *Journal of Cognitive Neuroscience*, 14(1):79–94.

Tyler, L. K. and Cobb, H. (1987). Processing bound grammatical morphemes in context: The case of an aphasic patient. *Language and Cognitive Processes*, 2(3-4):245–262.

Tyler, L. K. and Marslen-Wilson, W. (1986). The effects of context on the recognition of polymorphemic words. *Journal of Memory and Language*, 25(6):741–752.

Tyler, L. K., Marslen-Wilson, W., Rentoul, J., and 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.

Ussishkin, A. (1999). The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output–output correspondence. *Phonology*, 16(03):401–442.

Ussishkin, A. (2005). A fixed prosodic theory of nonconcatenative templaticmorphology. *Natural Language & Linguistic Theory*, 23(1):169–218.

Ussishkin, A., Dawson, C. R., Wedel, A., and Schluter, K. (2015). Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience*, 30(9):1096–1115.

Vannest, J., Newport, E. L., Newman, A. J., and Bavelier, D. (2011). Interplay between morphology and frequency in lexical access: The case of the base frequency effect. *Brain research*, 1373:144–159.

Versteegh, K. (1997). *The Arabic Linguistic Tradition. Landmarks in Linguistic Thought III*. London-New York: Routledge.

Warschauer, M. (2002). The internet and linguistic pluralism. *Silicon literacies: Communication, innovation and education in the electronic age*, pages 62–74.

Watson, J. (2007). *The Phonology and Morphology of Arabic*. The Phonology of the World's Languages. OUP Oxford.

Whaley, C. (1978). Wordnonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17(2):143–154.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.

Wray, S. and Ali, A. (2015). Crowdsource a little to label a lot: Labeling a speech corpus of Dialectal Arabic. In *Interspeech*.

Wray, S., Boutz, J., Novak, V., and Buckwalter, T. (2014). Authenticating and supplementing a yemeni arabic dictionary using a web corpus. Presented at American Association of Corpus Linguistics, Flagstaff, Arizona.

Wray, S., Mubarak, H., and Ali, A. (2015). Best Practices for Crowdsourcing Dialectal Arabic Speech Transcription. In *Proceedings of Workshop on Arabic Natural Language Processing*.

Wurm, L. H. (2000). Auditory processing of polymorphemic pseudowords. *Journal of Memory and Language*, 42(2):255–271.

Yates, M., Locker, L., and Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, 11(3):452–457.

Zaghouani, W. and Dukes, K. (2014). Can Crowdsourcing be used for Effective Annotation of Arabic? In *LREC*, pages 224–228.

Zaidan, O. F. and Callison-Burch, C. (2011a). Crowdsourcing translation: Professional quality from non-professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1220–1229. Association for Computational Linguistics.

Zaidan, O. F. and Callison-Burch, C. (2011b). The Arabic Online Commentary Dataset: an annotated dataset of informal Arabic with high dialectal content. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 37–41. Association for Computational Linguistics.

Zribi, I., Boujelbane, R., Masmoudi, A., Ellouze, M., Belguith, L., and Habash, N. (2014). A Conventional Orthography for Tunisian Arabic. In *Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland.*