Orthographic and Root Frequency Effects in Arabic: Evidence from Eye Movements and Lexical Decision

Ehab W. Hermena[1], Simon P. Liversedge[2], Sana Bouamama[3], and Denis Drieghe[3]

[1]Zayed University, UAE

[2]University of Central Lancashire, UK

[3]University of Southampton, UK

Author Note

Ehab W. Hermena, Cognition and Neuroscience Research Laboratory, Department of Psychology, College of Natural and Health Sciences, Zayed University, Dubai, UAE.

Simon P. Liversedge, School of Psychology, University of Central Lancashire, Preston, Lancashire, PR1 2HE

Sana Bouamama and Denis Drieghe, Centre for Vision and Cognition, Psychology, University of Southampton, Southampton, England.

Correspondence regarding this article should be addressed to Ehab W. Hermena, Cognition and Neuroscience Research Laboratory, Department of Psychology, College of Natural and Health Sciences, Zayed University, Academic City, P.O. Box 19282, Dubai, UAE, E-mail: Ehab.Hermena@zu.ac.ae

This research was supported by Leverhulme Trust Research Grant RPG-2013-205. The first author was partially supported by Zayed University Grant R17010.

Abstract

One of the most studied and robust effects in the reading literature is that of word frequency. Semitic words (e.g., in Arabic or Hebrew) contain roots that indicate the core meaning to which the word belongs. The effects of the frequency of these roots on reading as measured by eye movements is much less understood. In a series of experiments, we investigated and replicated traditional word frequency effects in Arabic: Eye movement measures showed the expected facilitation for high- over low-frequency target words embedded in sentences (Experiment 1). The same was found in response time and accuracy in a lexical decision task (Experiment 3a). Using target words that were matched on overall orthographic frequency and other important variables, but that contained either high- or low-frequency roots, we found no significant influence of root frequency on eye movement measures during sentence reading (Experiment 2). Using the same target words in a lexical decision task (Experiment 3b), we replicated the absence of root frequency effects on real Arabic word processing. At first glance, the results may not appear to be in line with theoretical accounts that postulate early morphological decomposition and root identification when processing Semitic words. However, these results are compatible with accounts where morphological decomposition does occur but is followed by re-combination, and under certain conditions re-combination costs can eliminate or even reverse root frequency effects.

Key words: Reading Arabic; Eye Movements; frequency effects; Semitic morphology; Lexical decision

One of the most well documented effects on eye movement control during reading is that of word frequency. Numerous investigations reported and replicated word frequency effects whereby words that occur and are encountered more frequently in a language attract shorter and fewer fixations, and more skipping, compared to words that occur in the language less frequently (see e.g., Inhoff & Rayner, 1986; Hyönä, 2011; Juhasz & Pollatsek, 2011; Rayner, 1998; 2009 for reviews). Word frequency effects are typically explained as a function of repeated encounters with a word affecting the speed with which the representations of this word are accessed and activated.

The experiments reported here investigate word frequency effects in Arabic. Arabic is a Semitic language that is read from right to left. In the first experiment the focus was on the orthographic frequency of the whole word, and how these influence fixation durations and other measures of eye movement control during sentence reading. In the second experiment the focus of investigation was whether the frequency of Arabic roots embedded within words influences eye movements behavior. Previous investigations of the processing of compound words, as well as prefixed and suffixed words in Finnish and in English suggested that readers engage in decomposing the morphological units of these words during word identification and reading, particularly for longer words (for a review, see Juhasz & Pollatsek, 2011). Furthermore, the frequency of these morphological units influences the length of fixations the word receives, particularly early fixations (e.g., Hyönä, Bertram & Pollatsek, 2004). Arabic however features Semitic morphology where the main morphological unit, namely the word root morpheme, is not located as an uninterrupted unit in the word (e.g., a unit that is flanked by a prefix, suffix, or both in English such as *order* in *preordered*; or as a part of a compound word, known as a lexeme, e.g., the words *black* or *bird* in the compound *blackbird*). Rather, Arabic morphology is non-concatenated, where root letters are typically diffuse within the word and can be interrupted by inserting letters

from the word form morpheme between the root letters (so-called infixes, e.g., مكتوب /mktub/ *is written*, where the root كتب /ktb/ is interrupted by the letter و /u/ of the form morpheme / مـ - و - / /m_ _ u _/, see e.g., Boudelaa & Marslen-Wilson, 2001; 2005; Schulz, 2004). In Semitic languages such as Arabic and Hebrew, the root morpheme indicates the main semantic family to which the word belongs (e.g., in the example above, the root كتب /ktb/ refers to writing-related meanings), whereas the form morpheme provides the detailed phonological representation, syntactic case, meaning, and gender, number, and tense inflections that are necessary for complete and accurate word identification (e.g., Boudelaa & Marslen-Wilson, 2005).

Roots play a very important role at an early stage in Semitic word identification. Previous investigations in Hebrew single word naming and lexical decision repeatedly suggested that the lexical organization of Semitic words (words that feature Semitic morphology to be precise) is root-, and not orthography-based (e.g., Deutsch, Frost & Forster, 1998; Deutsch, Frost, Pollatsek, & Rayner, 2000; Frost, Deutsch, & Forster, 2000; Frost, Deutsch, Gilboa, Tannenbaum, & Marslen-Wilson, 2000; Frost, Forster, & Deutsch, 1997; Frost, Kugler, Deutsch, & Forster, 2005; see also Frost, 2009; 2012 for review). The findings from these investigations repeatedly point to facilitation (typically shortening of response time) for word identification or lexical decision when root-sharing primes were provided, compared to when orthographically similar primes were provided. Similarly, Deutsch, Frost, Peleg, Pollatsek, and Rayner (2003) reported decreases in fixation durations (more specifically, in the measure of gaze duration which sums the fixation durations of first pass fixations on a target word) following the presentation of root-sharing previews compared to orthographically similar previews during sentence reading in Hebrew. Other supporting evidence was obtained in Arabic where statistically reliable facilitation in lexical decision was observed from root priming compared to form-related and orthographically-related

primes (Boudelaa & Marslen-Wilson, 2005). Boudelaa and Marslen-Wilson (2004) found that benefit from primes that shared root information also occurred for so-called weak roots, where the 3-letter root comprises a vowel that may change in one of the word derivations, thus only two consonants are shared between a prime and target (e.g., the root وفق /wfq/ with the first letter being a vowel و, in the word pair اتفاق /itifaq/ *agreement*, and وافق /wafaqa/ *agreed*, where only the root consonants /fq/ are shared). This facilitation was also found in prime-target pairs that shared the same weak root and that were semantically distant (e.g., the root وجه /wgh/ with the first letter being a vowel و, in the word pair واجه /wajaha/ *confronted* and اتجاه /itijah/ *direction* or *destination*, where the meanings of the word pair vaguely share the idea of what is in front of one's face). Similarly, other evidence from cross-modal priming further illustrated that the contribution of root information to word identification is clearly not reducible to mere number of shared letters, phonology, or even to the semantic closeness of the prime-target pair, further supporting the idea that Semitic lexicon organization is root-based (Boudelaa & Marslen-Wilson, 2015, see also Boudelaa, 2014; Boudelaa, Pulvermüller, Hauk, Shtyrov, & Marslen-Wilson, 2009; also Prunet, Béland, & Idrissi, 2000 for supporting evidence from a case study of an aphasic patient; and Gwilliams & Marantz, 2015 for supporting evidence from auditory processing of spoken Arabic). Indeed, the root-based organization of lexical entries has influenced print practices for centuries: Arabic dictionaries are known to order the word entries by roots, rather than by orthographic representations. For instance, to look up the word مكتوب /mktub/ the reader must find the root كتب /ktb/ first, and under it all derived forms of the root are listed.

If Arabic words that feature Semitic morphology are indeed lexically organized on the basis of roots, rather than on the basis of orthography, then, arguably, root frequency may influence the speed of word identification during reading. Thus, the question motivating Experiment 2 is whether the frequency of the root in Arabic words influences eye movement

behavior when the overall word frequency is held constant. To our knowledge, this is the first direct investigation of this question in Arabic reading. However, for consistency with findings across other languages, we will first replicate the traditional orthographic frequency effect in Arabic by comparing eye movement measures on target words that have high orthographic and root frequencies with target words that have low orthographic and root frequencies. Note that using the Aralex database (Boudelaa & Marslen-Wilson, 2010), it was not possible to find enough words that have both high orthographic frequency and low root frequency, so we were unable to match root frequency between the two orthographic frequency conditions, and as a result we cannot have a straightforward 2 orthographic frequency (low – high) × 2 root frequency (low – high) design.

Experiment 1

In the first experiment, we aimed to replicate the classic orthographic frequency effect on eye movements during reading that is widely reported in reading research (see above) in Arabic sentences. We expected to replicate this effect whereby Arabic words of high orthographic frequency will attract shorter fixation durations compared to words that are of low orthographic frequency. The analyses conducted also included the pre-target region to investigate possible effects of the target word frequency on fixation durations on the previous word (so-called parafoveal-on-foveal effects).

Method

Participants

Forty-two adult native Arabic speakers were paid £15 for participation in the eye tracking procedure. Only participants who were born in Arabic speaking countries, with Arabic as their first language, were classed as native readers and were allowed to participate. All participants were UK residents or visitors. The participants (24 females) had a mean age of 31 years (SD = 8.9, range = 18 – 54). All participants had normal or corrected to normal vision, and all reported being able to clearly see the words on the screen during a practice block. The majority of participants spoke and read English as a second language. All participants read Arabic text regularly (daily or weekly) and they were naïve as to the exact purpose of the experiments.

For the stimuli norming tasks detailed below, we used Amazon Mechanical Turk (AMT) participants, who did not take part in the eye tracking procedure. The exact number of AMT participants participating in each task is detailed below. AMT participants were paid £10-15, depending on the number of tasks in which they participated. AMT participants' Arabic reading skills were tested in a number of 'quality check' tasks embedded in the norming procedures (e.g., providing accurate definitions of the target words, and placing the target words in original, grammatically sound, sentences). Additionally, all tasks were time capped and so only highly skilled readers of Arabic were able to complete the work in the time allowed. Data from AMT participants whose work did not pass the quality checks were not included in the norming, and additional AMT participants were recruited to replace them.

Reading skill screening for participants in the eye tracking procedure.

Two reading tasks were performed by all participants prior to the experiment in order to screen their proficiency in reading Arabic. Participants performed a word reading aloud task (82 printed words) followed by a sentence reading aloud task (5 sentences including 42

words) presented on the computer screen. All participants were highly accurate both in word and sentence reading (mean percentage of words read accurately = 99.2%, SD = 1, range = 96.3 – 100%).

Stimuli

Thirty sets of two target words (total 60 words) of high and low orthographic frequency were selected from the Aralex database (Boudelaa & Marslen-Wilson, 2010). The target words were embedded in frame sentences that were identical in 19 of the stimuli sets (see Figure 1). For the remaining sets the frame sentences were identical only until the target word. After the target word, the remaining portion of the sentence differed between the conditions to suit the different target words. In each set, the target word pairs (high and low orthographic frequency) contained the same number of letters. Target words were always embedded near the middle of the sentence. Appendix 1 contains the stimuli sentences and lists the syntactic cases of all target words used. On average, the target words were 8.6 letters long (SD = 0.8, range = 8 – 11 letters). High-frequency words had an average orthographic frequency of 248.3 counts per million in Aralex (SD = 149, range = 100.8 – 680.5 counts per million). Low-frequency words had average orthographic frequency of 0.9 counts per million in Aralex (SD = 2.4, range = 0.2 – 9.7 counts per million). The difference in average log-transformed orthographic frequency was statistically significant $t(58) = 29.8$, $p < .001$. Additionally, high orthographic frequency words contained roots that had an average of 2950.8 counts per million (SD = 1824.8, range = 996.2 – 6902.5), whereas low orthographic frequency words contained roots that had an average of 580.6 counts per million (SD = 1074.7, range = 3.12 – 3934.1). Thus, high orthographic frequency words featured roots that were also of a significantly higher frequency than the roots in the low orthographic frequency

words (difference in log-transformed frequency counts was significant: $t(58) = 7.2$, $p < .001$).

<Insert Figure 1 about here>

All sentences were written and displayed on a single line and in natural cursive script. The text was rendered in Traditional Arabic font, size 18 (in size roughly equivalent to English text in Times New Roman font size 14).

Stimuli matching and norming.

Arabic is typically printed in proportional fonts with letters naturally varying in size, thus words that contain the same number of letters may vary in their spatial extent, or the amount of horizontal space the word occupies (see Hermena, Liversedge, & Drieghe, 2016). To make sure this property did not result in a confound between the conditions, target words were matched on spatial extent. This was achieved through extending letter ligatures when necessary. Extending these ligatures would typically increase letters' spatial extent minimally (by a pixel or two) so that both words in a stimulus set would have the exact same spatial extent of the largest one.

We obtained 10 cloze predictability ratings for the target word in each sentence. In this procedure, 10 AMT participants were given sentences up to, but not including, the target word, and were asked to complete the sentence. None of the target words were produced by the AMT participants indicating that none of these words were predictable (i.e., the target was produced on zero occasions by the AMT participants).

Finally, we obtained ratings of sentence structure naturalness for all target sentences on a 7-point scale (1 = structure is highly unusual, 7 = structure is highly natural). 10 ratings

per sentence were obtained from 10 AMT raters, and these indicated that sentence structure for all stimuli in all conditions was highly natural: Sentences containing high and low-frequency target words had average ratings of 5.98 (SD = 0.81, range = 5.3 – 6.5), and 5.59 (SD = 0.80, range = 5.4 – 6.4), respectively. There was no significant difference between the naturalness ratings of the two conditions ($t < 1$).

Apparatus

An SR Research Eyelink 1000 eye tracker was used to record participants' eye movements during reading. Viewing was binocular, but eye movements were recorded from the right eye only. The eye tracker sampling rate was set at 1000 Hz. The eye tracker was interfaced with a Dell Precision 390 computer, and with a 20 inch ViewSonic Professional Series *P227f* CRT monitor. Monitor resolution was set at 1024 × 768 pixels. The participants leaned on a headrest to reduce head movements. The words were in black on a light grey background. The display was 73 cm from the participants, and at this distance, on average, 2.3 characters equaled 1° of visual angle.

The participants used a VPixx RESPONSEPixx VP-BB-1 button box to enter their responses to comprehension questions and to terminate trials after reading the sentences.

Design

The orthographic frequency of the target words was the within-participants independent variable. Sentences containing these targets were counterbalanced, and presented in random order. Thus, participants saw only one sentence out of each set, and an equal number of target stimuli from both frequency conditions.

Procedure

This experiment was approved by the University of Southampton Ethics Committee. At the beginning of the testing session, participants were given instructions for the experiment. Consenting participants subsequently read aloud the words and the sentences for the reading skill screening task. This was followed by the eye tracking procedure.

The eye tracker was calibrated using a horizontal 3-point calibration at the beginning of the experiment, and the calibration was validated. Calibration accuracy was always < 0.25°, otherwise calibration and validation were repeated. Prior to the onset of the target sentence, a circular fixation target (diameter = 1°) appeared on the screen in the location of the first character of the sentence. When the tracker registered a stable fixation on the circle, the sentence was presented.

The participants were told to read silently, and that they would periodically be required to use the button box to provide a yes/no answer to the questions that followed around one-third of the sentences. Participants were allowed to take breaks, following which the tracker was re-calibrated. The testing session, including the reading skill screening tasks, the eye tracking procedure, and breaks lasted around 60 minutes, depending on how many breaks a participant took.

Eye tracking data for Experiments 1 and 2 were collected in one testing session, along with the stimuli from another, unrelated experiment. Thus, the sentences from the different experiments acted as filler items for each other. In total, participants read 96 sentences (30 from Experiment 1 + 30 from Experiment 2 + 26 from the unrelated Experiment 3 + 10 practice sentences).

Results

For all reported analyses, fixations with durations shorter than 80 ms, or longer than 800 ms were removed. However, fixations shorter than 80 ms that were located within 10 pixels or less from another longer fixation, were merged with the longer fixation. Along with removing trials in which blinks occurred, this resulted in removing approximately 0.6% of all data points. Furthermore, for each of the fixation duration measures, we removed data points ±2.5 standard deviations away from the mean fixation duration per participant within the specific condition as outliers.

Three participants were excluded from the analyses given that their sentence comprehension scores fell below 80%. Thus, the reported results are based on data collected from 39 participants. These 39 participants had an average sentence comprehension score of 94% (SD = 4, range = 80 – 100%). There were no differences between the accuracy scores across the conditions ($t < 1$).

We report a number of eye movement measures for the target word region. The first measure is *word skipping probability* (the probability that the target word was not fixated during first pass reading). We also report *first fixation duration* (the duration of the first fixation in first pass reading on the target word, regardless of the number of fixations the word received overall); *single fixation duration* (the duration of the fixation on the target in first pass reading in instances where the target received exactly one fixation during sentence reading); *gaze duration* (the sum of fixation durations the target word received during first pass reading and before exiting the target word to go forward or backwards in the text); and *go past time* (the sum of all fixation durations made from entering the region of interest until exiting this region forward). Finally, we also report *first pass fixation count* (the total number of fixations the word received during first pass reading).

In addition, we also report the duration of the last fixation of first pass reading and gaze duration on the pre-target word to learn whether there were any so-called parafoveal-on-foveal effects associated with the orthographic frequency of the target words (for a review see Drieghe, 2011).

We used the *lme4* package (version 1.1-16, Bates, Maechler, & Bolker, 2015) within the R environment for statistical computing (R-Core Development Team, 2016) to run linear mixed models (LMMs). Target word frequency (two levels: high vs. low) was the fixed factor for each model. Subjects and items were treated as random variables. Unless indicated below, all models used for fixation duration and fixation count measures contained the full random structure (e.g., Barr, Levy, Scheepers, & Tily, 2013) that included random slopes for the main effects and their interactions. For the measure of word skipping we used logistic generalized linear mixed models (GLMMs). If a model containing the full random structure failed to converge, it was systematically trimmed until it converged, first by removing correlations between random effects, and if necessary also by removing their interactions. All findings reported here are thus from successfully converging models. We performed log transformation of the fixation durations to reduce distribution skewing (Baayen, Davidson, & Bates, 2008). For each eye movement measure we report beta values (b), standard error (SE), *t* statistic for fixation durations and count measures, and *z* statistic for skipping probability. As a *t* distribution with a high degree of freedom approaches the *z* distribution, absolute *t* values higher than 1.96 can be considered significant at $p < .05$. Table 1 contains the descriptive statistics for all reported measures. All descriptive statistics (means and standard deviations) reported in this experiment, and the rest of the experiments in the current paper, were calculated across participants.

<Insert Table 1 about here>

Pre-Target Word Analysis

At the pre-target region, removing outliers from the last fixation duration of first pass reading resulted in removing 1.2% of data points, and 1.4% from gaze duration. There were no significant differences between the two conditions in the last fixation duration of first pass reading (b = 0.029, SE = 0.023, t = 1.26), or in gaze duration (t < 1).

Target Word Analysis

Removing fixation duration outliers resulted in removing 1.5% data points from first fixation duration, 0.4% from single fixation duration, 1.8% from gaze duration, and 3.3% from go past time.

As can be seen in Table 1, low-frequency words were slightly more likely to be skipped compared to high-frequency words, however the difference was not significant (b = 0.443, SE = 0.367, z = 1.21, p >.20)[1]. It is notable that, overall, target word skipping was quite rare. Furthermore, and as expected, compared to low-frequency targets, high-frequency targets received a significantly shorter first fixation duration (b = 0.104, SE = 0.024, t = 4.38), single fixation duration (b = 0.121, SE = 0.033, t = 3.07), gaze duration (b = 0.259, SE = 0.034, t = 7.70), and go past time (b = 0.312, SE = 0.039, t = 7.91). High-frequency words also attracted significantly fewer first pass fixations compared to low-frequency words (b = 0.272, SE = 0.065, t = 4.21).

[1] The model with full random structure failed to converge and was thus trimmed. The converging version of the model was: glmer (dependent_variable ~ frequency_condition + (1 | participant) + (1 | stimulus_item), data = data_file, family = binomial).

Discussion

The results obtained replicate previous findings for word frequency effects in other languages. Arabic words of high orthographic frequency attracted shorter fixation durations in eye movement measures that are associated with early (first and single fixation durations, and gaze durations) as well as late (go past) processing. The results also indicated that the orthographic frequency of the target words did not influence processing time on pre-target interest areas. In other words, fixation duration measures suggest that there were no parafoveal-on-foveal effects for word frequency. There were no significant effects of word frequency of target word skipping.

As discussed above, word frequency effects in reading (and in other single word identification tasks) are robust findings that are widely reported and replicated. Word frequency effects are used as a benchmark for modelling of eye movement behavior in reading. As such, both families of eye movement control models, serial and parallel, successfully accommodate word frequency influences on eye movement control during reading. For instance, in E-Z Reader which postulates sequential attention allocation to words during reading, suggests that word frequency determines the average time needed to complete the familiarity check (L1, in combination with word predictability from previous context, see Reichle, 2011). On the other hand, in SWIFT, a parallel processing model which proposes gradient attention allocation during reading, word frequency effects are also present but not only for the currently fixated word as the model additionally accommodates successor effects (i.e., that fixation duration is modulated by the properties of the upcoming word, including its frequency) and lag effects (i.e., that word recognition continues to influence subsequent fixation durations after gaze position has shifted forward to the upcoming word,

see e.g., Engbert & Kliegl, 2011).

As this is the first report of word frequency effects on fixation durations during reading in Arabic, it would be interesting to consider this effect in some detail. While the eye movement control models discussed above successfully simulate or predict word frequency effects, they do not provide any explanation of why word frequency effects are obtained in the first place. This is a fundamental issue. According to Norris (2006), a number of cognitive modelling exercises in the word recognition literature do not answer this question either. In the explanations offered by some models such as the Logogen Model (Morton, 1969), or the search models family (e.g., Forster, 1976) the word frequency effect is treated as "an undesirable side effect of a suboptimal [word identification] mechanism" (Norris, 2006, p. 329). The essence of such explanations is that in that "suboptimal mechanism" a portion of words in the language, namely, those that occur less frequently, are disadvantaged even though they were encountered and learnt previously. By contrast, Norris (2006) offers a different account, in the Bayesian Reader model, that assumes that the word identification system actually functions optimally. In this model, word frequency effects occur because a word identification system that is optimally adapted to the linguistic environment in which it operates would by default identify more frequent words faster than less frequent ones (see also Norris & Kinoshita, 2008). As such, the Bayesian Reader model (Norris, 2006) assumes that when performing word identification, an ideal observer cannot simply match perceptual input (print) to all stored lexical entries (words), with each entry requiring the same amount of processing to be retrieved. Indeed, had this been the case, no word frequency effects are to be expected. Rather, the ideal observer takes into account the prior probabilities of the word occurrence, thus, inevitably, that observer would be influenced by how frequent the word appears in the particular language. Note that taking into account words' frequency of occurrence in a language is suggested to be a result of system optimization, and not because

the system sub-optimally functions when attempting to match perceptual input to a subset of the previously learnt and stored entries (namely, the subset of entries that are encountered less frequently in the language). This subtle point is perhaps the main difference between this account, and the accounts proposed by the models mentioned above. So, combining perceptual information with prior probability allows the observer to perform word identification, whether in reading or other tasks such as lexical decision in a way that maximizes performance speed and accuracy, and minimizes misidentification that could lead to erroneous response (e.g., in lexical decision), or to building inaccurate representations of the text during reading. Specifically, the probability of observing the perceptual input *I*, given that the word *W* has been presented is captured by term *P(I|W)*. Each time a word is encountered, the recognizer is able to learn and update that probability. Finally, in dealing with any new perceptual input, the system 'looks up' this probability *P(I|W)* in order to generate the desired response. Thus, optimal word identification system functioning produces, and replicates, word frequency effects.

Experiment 2

As explained above, a great deal of evidence emerging from studies of Hebrew and Arabic word processing suggests that the root morphemes in these words play a key role, not only in word identification, but also in lexical organization. In this experiment, we investigate whether high root frequency results in processing facilitation during sentence reading. Specifically, would words that contain a high-frequency root attract shorter fixation durations compared to words that contain low-frequency roots? The two sets of words are matched on length (number of letters), spatial extent, predictability from previous context, and notably on whole word orthographic frequency. If the words containing high-frequency

roots attract shorter fixation durations compared to words with low-frequency roots, this would be a very interesting finding that further illustrates the important role of root morphemes in word processing. Such results would further support the idea that the organization of lexical representation in Arabic is root-based, and as such: (a) More frequently encountered roots are faster to activate and easier to process compared to less frequently encountered roots; and importantly, (b) root frequency influences the processing time (fixation durations) required for the identification of the words containing these roots, similar to the way that orthographic frequency influences processing time in other languages where lexical organization is orthography-based (see Frost, 2012). Such findings would also complement previous findings from single word tasks in Arabic and Hebrew (e.g., primed lexical decision, see discussion above) where primes that activate root representations shared with targets result in facilitation (faster responses) to these targets.

Additionally, we suggest that obtaining an effect of root frequency on target word identification during reading would be predicted from the dual route model for processing Semitic words that was put forward by Frost et al. (1997). In this model, an obligatory morphological decomposition and root identification route was suggested to influence letter string processing at early stages, in combination with a whole-word processing route. Such a model could account for the robust findings that clearly suggest that Semitic language readers are sensitive to root information presented as primes (see also Bentin & Frost, 1995). It follows that if Arabic readers similarly decompose Arabic words into morphological units, high frequency roots would have a processing advantage compared to roots of lower frequency.

Similar to Experiment 1, eye movement measures on pre-target words were also analyzed to establish whether processing Arabic words with high or low root frequencies results in any parafoveal-on-foveal effects.

Method

The participants, apparatus, and procedure of this experiment are identical to Experiment 1. As explained above, collecting data for both experiments took place in the same session with the stimuli of both experiments, as well as a third unrelated experiment, acting as filler items for each other.

Stimuli

Thirty sets of two target words (total 60 words) of high and low word root frequency were selected from the Aralex database (Boudelaa & Marslen-Wilson, 2010). The target words were embedded in frame sentences that were identical in 18 of the stimuli sets (see Figure 2). For the remaining sets the frame sentences were identical only until the target word. In each set, the target word pairs (high and low root frequency) contained the same number of letters. Half the sets contained 6-letter target word pairs, and the other half 7-letter word pairs. The majority of target word sets contained 3-letter roots with only 2 sets containing 4-letter roots (both were in the group of the 6-letter words). This selection is representative of Arabic words where the majority of roots are 3-letters long (Haywood & Nahmad, 1965; Schulz, 2004; see also Buckwalter & Parkinson, 2011). In each set, the target word pair contained the same number of root letters. Target words were always embedded near the middle of the sentence. Appendix 2 contains the stimuli sentences and lists the syntactic cases of all target words used in each sentence. High-frequency roots had an average of 4959.8 counts per million in Aralex (SD = 6286.7, range = 273.8 – 31507.5 counts per million). By contrast, low-frequency roots had an average of 20.6 counts per million (SD

= 19.9, range = 0.2 – 65.0 counts per million). The difference in log-transformed root frequency counts between the two groups was statistically significant $t(58) = 15.95$, $p < .001$.

<Insert Figure 2 about here>

Both root frequency groups were matched on overall word orthographic frequency: High root frequency words had an average orthographic frequency of 1.45 counts per million in Aralex (SD = 2.14, range = 0.18 – 8.19 counts per million); low root frequency words had average orthographic frequency of 1.23 counts per million in Aralex (SD = 1.53, range = 0.18 – 7.10 counts per million)[2]. The difference between the log-transformed orthographic frequencies of these two groups was not statistically significant ($t < 1$). As mentioned above, it was not possible to find enough words with high orthographic frequency and low root frequency to construct a fully crossed design.

As with Experiment 1, all sentences were written and displayed on a single line and in natural cursive script. The text was rendered in Traditional Arabic font, size 18.

[2] In all reported experiments, root token frequencies, not type frequencies were the basis on which stimuli selection was performed. Root type frequency is an interesting variable given its potential influence on readers' performance (see e.g., Boudelaa & Marslen-Wilson, 2011). We did not include root type frequency in our manipulations or discussion as it falls outside the remit of our a priori research questions. Rather, our stimuli selection preserved the relationship between root type and token frequencies as present in the Aralex database (Boudelaa & Marslen-Wilson, 2010). Specifically, based on all 142,162 accessible root entries in Aralex, root type and token frequencies are strongly and positively correlated (log transformed type and token frequencies have a correlation coefficient of $r = 0.72$, $p < .001$). In our selected stimuli for all reported experiments, this relationship between root type and token frequencies was preserved ($r = 0.81$, $p < .001$ for all stimuli; $r = 0.60$, $p < .001$ for Exp. 1; $r = 0.82$, $p < .001$ for Exp. 2; real roots in Exp. 3a have the same properties as Exp. 1; $r = 0.84$, $p < .001$ for Exp. 3b; all frequency counts log transformed). We wish to thank an anonymous reviewer for alerting us to the relevance of including this information.

Stimuli matching and norming.

Target words were matched on spatial extent in a manner identical to Experiment 1. Similarly, none of the target words were predictable from the pre-target context based on 10 cloze predictability ratings obtained for each sentence stem provided by AMT participants.

Finally, we obtained ratings of sentence structure naturalness for all target sentences on a 7-point scale (1 = structure is highly unusual, 7 = structure is highly natural). 10 ratings per sentence were obtained from 10 AMT raters, and these indicated that sentence structure for all stimuli in all conditions was highly natural: Sentences containing high and low root frequency target words had average ratings of 5.80 (SD = 0.71, range = 5.4 – 6.6), and 5.88 (SD = 0.72, range = 5.3 – 6.6), respectively. There was no significant difference between the naturalness ratings between the two conditions ($t < 1$).

Results

Data cleaning criteria and procedure were identical to what is described in Experiment 1, and resulted in removing approximately 1.1% of all data points. No participants were excluded on the basis of sentence reading comprehension given that all scores were above 80% in this experiment (sentence comprehension scores were analyzed separately for Experiments 1 and 2). Thus, the analyses reported are based on the data from all 42 participants. On average participants had a comprehension score of 94% (SD = 5.1, range = 81 – 100%). There were no differences between the accuracy scores across the root frequency conditions ($t < 1$).

We report the same eye movement measures reported in Experiment 1 for the target

and pre-target regions. The linear mixed models used to analyze the data were specified in a manner similar to what is described in Experiment 1, with the exception that target word root frequency (two levels: high vs. low) was the fixed variable for each model. Unless indicated, all LMM and GLMM models used contained full random structures and successfully converged. Table 2 contains the descriptive statistics for all reported measures for Experiment 2.

<Insert Table 2 about here>

Pre-Target Word Analysis

At the pre-target region, removing outliers from last fixation duration of first pass reading resulted in removing 0.9% of data points, and 1.9% for gaze duration. See Table 2 for descriptive statistics for eye movement measures at the pre-target region. For the last fixation duration of first pass reading, the difference between the two conditions was small and not statistically significant ($t < 1$). Similarly, the difference between the two conditions was not significant for the gaze duration measure ($t < 1$)[3].

Target Word Analysis

Removing fixation duration outliers resulted in removing 1% data points from first fixation duration, 0.4% data points from single fixation duration, 0.9% from gaze duration,

[3] For both these measures, the models with full random structure resulted in random effects correlations of 1 or -1 indicating over-parameterization. The random structures of these models were thus trimmed. The models used were: lmer (dependent_variable ~ frequency_condition + (1 | participant) + (1 | stimulus_item), data = data_file).

and 2.9% of go past time.

As can be seen in Table 2, the difference between target word root frequency conditions in the measure of word skipping was negligible and not statistically significant ($z < 1$)[4]. Similarly, the differences between the two root frequency conditions were not statistically significant for first fixation or single fixation durations, gaze duration[5], go past time, or first pass fixation count[6] (all ts < 1).

An additional Bayesian Factor (BF) analysis was performed using the BayesFactor package (version 0.9.12-2, Morey & Rouder, 2015) for R (R-Core Development Team, 2016) and used the default scale value (0.5) for the Cauchy priors on effect size and 100,000 Monte Carlo iterations. Contrary to traditional null-hypothesis testing Bayesian statistics allow us to quantify the amount of evidence the data provide for either the null hypothesis or the alternative hypothesis. Applied to our current experiment it allows us to compare the amount of evidence for a model that did, or did not, include root frequency as a predictor. A low BF (<1) would indicate evidence for the simpler model, a high BF (>1) evidence for a model that does include root frequency. The BF was calculated for all reported dependent variables. For the pre-target word region, the BF analyses indicated what can be classed as strong evidence

[4] The model with full random structure failed to converge and was thus trimmed. The converging version of the model was: glmer (dependent_variable ~ frequency_condition + (1 | participant) + (1 + frequency_condition | stimulus_item), data = data_file, family = binomial).

[5] For first and single fixation durations, and gaze duration, the models with full random structure resulted in random effects correlations of 1 or -1 indicating over-parameterization. The random structures of these models were thus trimmed. The models used were: lmer (dependent_variable ~ frequency_condition + (1 | participant) + (1 | stimulus_item), data = data_file).

[6] For first pass fixation count the model with full random structure resulted in random effects correlations of 1 indicating over-parameterization. The random structures of this models were thus trimmed. The model used was: lmer (dependent_variable ~ frequency_condition + (1 | participant) + (1 + frequency_condition | stimulus_item), data = data_file).

for the absence of root frequency effects in the measure of last fixation duration in first pass reading (BF = 0.08; a BF smaller than 0.33 is usually considered to constitute substantial evidence for the null effect, and a BF smaller than 0.1 strong evidence), and substantial evidence for this null result in the measure of gaze duration (BF = 0.12). Similarly, BF analyses showed evidence for the absence of root frequency effects in all reported measures at the target word region (skipping: BF = 0.22, substantial; first pass fixation count, BF = 0.12, substantial; first fixation duration: BF = 0.08, strong; single fixation duration: BF = 0.10, substantial; gaze duration: BF = 0.07, substantial; and go past: BF = 0.13, substantial). BF values indicating substantial or stronger support for null or alternative hypotheses are considered sufficient indicators that the data set does not lack sensitivity, or power, and that the null hypothesis (in the current results) is well-supported (see e.g., Dienes, 2014; Wetzels, Matzke, Lee, Rouder, Iverson, & Wagenmakers, 2011)[7].

Discussion

The results showed no difference in any of the eye movement measures as a function of root frequency. As discussed above, if readers utilize an obligatory morphological decomposition and root identification route, then we would have expected to obtain robust root frequency effects. However, we will argue that the current null results cannot be used as conclusive evidence against compulsory morphological decomposition and root identification (e.g., Frost et al., 1997). An alternative interpretation using the theoretical framework put

[7] To further examine whether these results can be attributable to lack of statistical power, we used the power analyses described by Westfall, Kenny, and Judd (2014). The analyses revealed that the number of stimuli items per cell in the current design (30), and current number of participants (42) would be sufficient to detect a moderate effect size $d = 0.5$ with power = 0.89. In other words, it is not at all likely that these null effects arose due to a lack of statistical power.

forward by Taft (2004) allows us to evaluate the viability of an obligatory morphological decomposition route in the light of the current results. To avoid repetition, however, we will detail this account in the General Discussion.

In order for this discussion to be comprehensive, we must consider findings in European languages where the frequency of morphological constituents (e.g., lexemes in compound words, like *color* in *colorscale*, a single word in Finnish, see Hyönä & Pollatsek, 1998) was found to influence fixation durations during silent reading, independently from overall word orthographic frequency. A possibility that warrants future investigation is that Arabic roots did not yield a frequency effect similar to lexemes in European languages because of a very important difference between the two types of word sub-components: Lexemes in European compounds represent an uninterrupted, isolable, portion of the word whereas the non-concatenated nature of Semitic morphology mean that roots in Arabic words are spread within the word, and are separated by letters from the word form morpheme. As a result, it is possible that processing Arabic roots and European lexemes may differ fundamentally in the way each influences eye movement control during silent reading, such that lexeme frequency effects on eye movement measures are more robust than those of Arabic roots frequencies. At this stage we can only offer speculations as to the exact mechanism by which the dispersal of root letters in Arabic words may eliminate the root frequency effects (unlike the documented frequency effects of isolable and unified lexemes in European words). One possibility is that the dispersal of root letters may result in slowing down root identification (compared to if the roots were present as unseparated letters). In turn, this slowing down of root identification may result from increased lateral inhibition (i.e., visual crowding, see Bouma, 1970, 1973; Drieghe, Brysbaert, & Desmet, 2005) that may slow down the identification of root letters, and is caused by the non-root letters that interrupt root unity. Further direct investigation of this issue may be necessary.

Another difference between words in Arabic and in European languages is that of word length (the number of letters a word contains). Findings from European languages (e.g., Finnish and English) show that for longer words (≥ 12 letters) lexeme frequency influences measures such as gaze duration, whereas shorter words (≤ 8 letters) show only effects of whole-word frequency (e.g., Bertram & Hyönä, 2003; Niswander-Klement & Pollatsek, 2006; see also Kuperman, Bertram, & Baayen, 2010 for comparable results in Dutch, but cf. Juhasz, 2008). Most Arabic words with Semitic roots (i.e., words that are not Arabized from other languages such as ديموقراطية or *democracy*, 10-letters in Arabic) are shorter than 10-letters long. This is the case even for words that include gender, number, and tense inflictions added to the root (e.g., ستتعاونان or *both [females] will cooperate*, total of 9 letters, root letters underlined, see Haywood & Nahmad, 1965; Schulz, 2004). Bertram and Hyönä (2003) reported that most words that were about 8-letters long attracted one fixation, whereas with longer Finnish compounds, more than one fixation was necessary. This meant that individual lexemes in longer words were most likely processed in different fixations, and the fixation durations on these lexemes reflected the frequency with which these lexemes occur in Finnish. In Arabic, by contrast, and given the relative shortness of Arabic words (in terms of number of letters; almost 40% of the words listed in Aralex are composed of ≤ 5 letters), and also the fact that root and form morpheme letters are dispersed, it is likely that the durations of each fixation made on these words reflects a mixture of processing both root and non-root letters. This may mean that fixation duration measures of Arabic word processing may not readily show a significant Semitic root frequency effect, despite the importance of these roots for lexical organization.

Given the results obtained for the root frequency manipulation, we decided to expand the investigation of word and root frequency effects (particularly the latter) in a less natural reading task. Recall that the findings relating to the central role of roots in word

identification in Semitic languages were mainly obtained from isolated word recognition tasks (e.g., lexical decision). In addition to replicating the reported effect of word frequency in Experiment 1, we aimed to further investigate the relationship between a root and the letter-string in which it is embedded in a way that would further explain why only word, but not root, frequency effects were obtained.

Experiments 3a and 3b

These two experiments further investigated word orthographic frequency effects (Experiment 3a) and root frequency effects (Experiment 3b) on lexical decision performance. In both experiments, we re-used the target words from Experiments 1 and 2 as the word items. This allowed for investigation of whether, using the same stimuli, the orthographic frequency effects obtained in the eye tracking experiment (Experiment 1), generalized to lexical decision performance.

For Experiment 3a, we expected that word frequency would influence lexical decision performance such that response times would be reduced for high- relative to low-frequency words. We also expected a similar effect in response accuracy.

Importantly, for Experiment 3b, using the same stimuli we used in silent reading (Experiment 2) in lexical decision allowed us to investigate whether the pattern of results reported above extends to lexical decision. The results reported above reflected effects of word orthographic frequency, while suggesting that the frequency of the roots these target words encompass does not significantly influence eye movement measures of reading. In a lexical decision task, using pseudo words that contain real roots, of either high or low frequency, allows us to investigate root frequency influence on letter string identification when these letter strings represent real lexical entries compared to when these strings are

novel. We can thus disentangle and quantify word orthographic and root frequency effects on letter string identification. This motivated the method we describe below for creating the pseudo words used in the lexical decision task. It also motivated including target letter string lexicality (real word vs. pseudo word) as a fixed variable in our statistical analyses, in addition to the variable of root frequency). The findings from this experiment would also be informative in evaluating the viability of a compulsory morphological decomposition and root identification route (e.g., Frost et al., 1997), and its influence on lexical decision.

Method

Participants

Forty-five adult native Arabic speakers were paid £15 for participation in these two experiments (as well as other, unrelated Arabic sentence reading experiments that were run simultaneously). None of the participants in these experiments took part in Experiments 1 or 2. The criteria for selecting native Arabic readers was operationalized as described in Experiment 1. The participants (22 females) had a mean age of 31 years (SD = 6.7, range = 19 – 50). All participants had normal or corrected vision, and all reported being able to clearly see the words on the screen during a practice block.

Reading skill screening for participants.

In order to screen the proficiency of reading in Arabic, three reading tasks were completed by all participants. Prior to the experiment participants performed a text reading aloud task (82 printed words) followed by a sentence reading aloud task (5 sentences including 42 words) presented on the computer screen. All participants were proficient with

100% accuracy. Subsequent to the eye tracking procedure, a standard digital voice recorder was used to record participants' voices when reading aloud a list of single words (36 words carrying Arabic diacritical marks which add vowels sounds to the letters). All 44 participants were highly accurate in word reading (mean percentage of words read accurately = 96.7%, SD = 4.0, range = 82% – 100%).

Stimuli

For Experiment 3a, we used the 30 sets of word pairs used in Experiment 1 as target words of high and low orthographic frequency. In addition, we created a set of 60 pseudo words. These pseudo words were paired with each of the high and low-frequency words such that the pseudo words contained the same number of letters, and the same number and pattern of morphemes as the real words. The pseudo words were built from pronounceable Arabic letter combinations. The roots contained in these pseudo words were composed of nonsense letter strings that did not correspond to any root entries in any of the major nine Arabic language dictionaries[8]. Thus, we ascertained that none of the pseudo words contained any real Arabic roots of either contemporary or archaic use.

For Experiment 3b, we used the 30 sets of word pairs used in Experiment 2 as target words of high and low root frequency (both matched on low orthographic frequency, see

[8] These are:

لسان العرب، مقاييس اللغة، الصحاح في اللغة، القاموس المحيط، العباب الزاخر، مختار الصحاح، المعجم الوسيط، تاج العروس، ومعجم اللغة العربية المعاصرة .

We used the electronic searchable versions of these dictionaries available at http://www.maajim.com and http://www.baheth.info.

details above). In addition, we created another 30 sets of pronounceable pseudo word pairs, half contained high-frequency real Arabic roots (average root count per million in Aralex = 4438.8, SD = 1593.5, range = 333.8 – 8294.2), and the other half low-frequency real Arabic roots (average root count per million = 0.7, SD = 0.2, range = 0.1 – 0.9). The difference between root frequency (log-transformed) in both pseudo word conditions was statistically significant ($t(58) = 62.7$, $p < .001$). The pseudo words in both these conditions were thus paired with the real words in both the high and low root frequency conditions such that each pseudo word matched the real word on number of letters and number and pattern of morphemes. This way we were able to orthogonally manipulate target lexicality (word or pseudo word) and root frequency (high or low). Appendix 3 contains all the target words and pseudo words used.

For both Experiments 3a and 3b, all words and pseudo words were displayed at the center of a computer screen in natural cursive script in Traditional Arabic font, size 18.

Apparatus

In both experiments, the lexical decision task was prepared using Experiment Builder (SR Research Ltd., Kanata, Ontario, Canada). The target letters strings were displayed at the center of a 20 inch ViewSonic Professional Series *P227f* CRT monitor and were viewed binocularly. Monitor resolution was set at 1024 × 768 pixels running at 120 Hz vertical refresh rate. A Dell Precision 390 computer handled the experimental display. The participants leaned on a headrest while viewing the targets. The targets were in black on a light grey background. The display was 73 cm from the participants.

The participants used a Dell SK-8511 computer keyboard to enter their responses to the lexical decision task (word: right ctrl / pseudo word: left ctrl).

Design

In Experiment 3a, the orthographic frequency of the target words, and lexicality of the letter strings were the within-participant independent variables. The stimuli were counterbalanced using a Latin square, and presented in random order. Thus, participants saw each target only once, and an equal number of high and low-frequency words, and an equal number of words and pseudo words in the testing session.

In Experiment 3b, 2 target lexicality (word, pseudo word) × 2 root frequency (low, high) were the within-participant independent variables. The stimuli were counterbalanced using a Latin square, and presented in random order. Thus, participants saw each target only once, and an equal number of words and pseudo words, and of high and low root frequency targets in the testing session.

In both experiments, button press (lexical decision) reaction time and accuracy were the dependent variables.

Procedure

Both experiments were approved by the University of Southampton Ethics Committee. At the beginning of the testing session, participants were given instructions for the experiments. Consenting participants subsequently read aloud the reading skill screening text before and after the lexical decision procedure, concluding the session with single word reading task.

The targets from both Experiments 3a and 3b were interleaved and presented at the same testing session. Each participant thus saw a grand total of 130 letter strings for lexical

decision: 60 targets from each experiment, plus 10 practice items. The procedure for presenting the target strings on the screen resembled the procedure for lexical decision used by Luke and Christianson (2011). Each trial began with a fixation circle at the center of the monitor, the exact location where the target string appeared. The fixation circle occupied the center of the screen for 300 ms. The circle was then replaced by the new target letter string. The target letter strings were displayed for a maximum of 3000 ms. Once the participant has responded by a button press, an 11-character mask ########### (equal in width to the widest target word) was displayed in the same location as the target, and remained for 200 ms, followed by a blank screen that was then displayed for 300 ms. The participants were then presented with the screen containing the fixation circle for the next trial.

The testing session, including participating in the other, unrelated sentence reading experiments and the reading skill screening tasks, lasted around 60 minutes.

Results

For both Experiments 3a and 3b, trials where reaction times shorter than 250 ms and longer than 1500 ms were removed (see e.g., Perea, Abu Mallouh, & Carreiras, 2010). Additionally, reaction time was analyzed only for trials where the participants' responses were accurate.

We used the *lme4* package (version 1.1-16, Bates et al., 2015) within the R environment for statistical computing (R-Core Development Team, 2016) to run linear mixed models (LMMs). For Experiment 3a, target condition (words with high orthographic frequency, words with low orthographic frequency, and pseudo words) was the within-participant fixed variable for each model. We pre-specified the words with high-frequency condition as the baseline to which we contrasted the other two conditions. Subsequently we

contrasted the baseline with low orthographic frequency and with the pseudo word conditions.

For Experiment 3b, the data were analyzed using Linear Mixed-effects Models (LMM) using the same lme4 package in R. Contrasts were specified as -.5/.5 and were used for the effects of target lexicality (word, pseudo word) and root frequency (low, high), such that the intercept corresponds to the grand mean and the fixed effects correspond to the main effect of the fixed factors.

Response time analyses for both experiments yielded very similar results when raw response times and log-transformed response times were analyzed. We thus report raw response time analyses for both experiments to preserve transparency. Furthermore, in the statistical models of both experiments subjects and items were treated as random variables. For the measure of button press accuracy, we used logistic generalized linear mixed models (GLMMs). All LMM and GLMM models contained full random structures that were trimmed (where indicated) following the procedure outlined above if failure to converge occurred. For both the reaction time and accuracy measures we report beta values (b), standard error (SE), *t* statistic for reaction time, and *z* statistic for accuracy. Tables 3 and 4 contain the descriptive statistics for Experiments 3a and 3b, respectively.

<Insert Tables 3 & 4 about here>

Experiment 3a. Removing trials with reaction times shorter than 250 ms and longer than 1500 ms resulted in removing around 15% of data points. Additionally, for reaction time analyses, 6% of data points were removed because of inaccurate responses.

Participants were significantly more accurate in performing lexical decision on high-frequency words compared to low-frequency words (b = 3.31, SE = 0.41, $z = 8.04$), and

compared to pseudo words (b = 2.94, SE = 0.41, z = 7.10) [9]. An additional contrast showed that participants were less accurate responding to low-frequency words compared to pseudo words (b = 0.418, SE = 0.154. z = 2.7). Similarly, participants responded significantly faster to high-frequency words compared to low-frequency words (b = 195.38, SE = 18.34, t = 10.65), and compared to pseudo words (b = 273.25, SE = 21.90, t = 12.48) [10]. Participants were also faster responding to low-frequency words compared to pseudo words (b = 133.74, SE = 20.67, t = 6.47).

Experiment 3b. Removing trials with reaction times shorter than 250 ms and longer than 1500 ms resulted in removing around 16% of data points. Additionally, for reaction time analyses, 10% of data points were removed because of inaccurate responses.

A significant effect for target lexicality on response accuracy was obtained (b = 0.51, SE = 0.11, z = 4.60). There was no main effect of root frequency on accuracy (z < 1)[11]. There was a significant interaction between target lexicality and root frequency (b = 1.12, SE = 0.22, z = 5.10, see Figure 3). For real words, response accuracy was significantly higher for high-frequency roots compared to low-frequency roots (z = 3.69), whereas the opposite pattern was obtained for pseudo words: Accuracy scores were significantly higher for low- compared to high-frequency roots (z = 3.45). Subsequent simple effects tests also revealed that for high-frequency roots, response accuracy was significantly lower for pseudo words

[9] The model with full random structure failed to converge and was thus trimmed. The converging version of the model was: glmer (dependent_variable ~ condition + (1 | participant) + (1 | stimulus_item), data = data_file, family = binomial).

[10] The model with full random variables failed to converge. The converging version was: lmer(dependent_variable ~ condition + (1 + condition | participant) + (0 + condition | stimulus_item), data = data_file)

[11] The model with full random structure failed to converge and was thus trimmed. The converging version of the model was: glmer(dependent_variable ~ item_lexicality * Root_Frequency + (1|pp) + (1|stim), data = data_file, family = "binomial")

compared to real words ($z = 6.64$), whereas for low-frequency roots, there was no difference between pseudo words and real words ($z < 1$).

<Insert Figure 3 about here>

For response time, there were significant main effects for both target lexicality (b = 131.54, SE = 22.60, $t = 5.82$), and root frequency (b = 19.44, SE = 9.45, $t = 2.06$). Importantly, these effects were qualified by a significant interaction between the two variables (b = 76.90, SE = 26.25, $t = 2.93$, see Figure 4). Subsequent simple effects tests revealed that there was no significant difference between response times for high- and low-frequency roots in real word targets ($t = 1.34$), whereas in pseudo words response times to high-frequency roots were significantly longer than for low-frequency roots ($t = 3.23$). Also, response times for pseudo words with high- and low-frequency roots were significantly longer compared to real words with high-frequency roots ($t = 6.67$) and low-frequency roots ($t = 3.48$).

Same as with Experiment 2, the Bayesian analysis was performed using the BayesFactor package (Morey & Rouder, 2015) for R (R-Core Development Team, 2016) to allow us to determine the extent to which our data can be used to conclude that there is no effect of root frequency in the experiment. The BF was calculated for response accuracy and latency, comparing models that included root frequency as a predictor variable to models that did not (collapsing across the lexicality variable). The BF analyses indicated substantial evidence for the absence of root frequency effects in both response accuracy (BF = 0.32), and response time (BF = 0.19). As explained above, BF values indicating substantial or stronger support for null or alternative hypotheses are considered sufficient indicators that the data set does not lack sensitivity, or power, and that the null hypothesis (in the current results) is well-

supported (Dienes, 2014; Wetzels et al., 2011)[12].

<Insert Figure 4 about here>

Discussion

In Experiment 3a, the results showed an orthographic frequency effect on lexical decision response time and accuracy rate. High-frequency words yielded the shortest reaction times, and the highest accuracy scores compared to low-frequency and pseudo words; and low-frequency words were responded to significantly faster compared to pseudo words. We were thus able to obtain consistent word frequency effects for Arabic words in eye movement measures during sentence reading, as well as in reaction time and response accuracy in lexical decision.

In Experiment 3b we obtained a significant effect of root frequency on response time in lexical decision. This significant effect was however qualified by a significant interaction with letter string lexicality—a variable that also had a significant main effect on response time and accuracy. The results show that there were no root frequency effects for real words on the measure of response time. Rather, only a small (4%) response accuracy advantage

[12] Although in Experiment 3b we were able to detect significant effects of root frequency, target string lexicality, and a significant interaction, we used the power analyses described by Westfall et al. (2014) once again to further examine whether the absence of root frequency effect on response time latency for real words (see Table 4) was due to lack of statistical power. Note also that in this respect, Experiment 3b results replicate the absence of significant root frequency effect on processing time (fixation durations) in the adequately-powered Experiment 2. The analyses revealed that for Experiment 3b, the number of stimuli items per cell in the current design (15), and current number of participants (45), the experiment could detect a moderate-to-high effect size $d = 0.57$ with adequate power = 0.8.

was found for high-frequency roots embedded in real words, relative to real words containing low-frequency roots. These results can thus be considered a replication of the results reported in Experiment 2: Root frequency did not have a significant effect on fixation durations in silent reading, nor lexical decision response latency, as both are measures of processing time.

By contrast, the analyses of lexical decision performance on pseudo words revealed a significant effect of root frequency, however not in the typical direction for frequency effects. The presence of high-frequency roots in pseudo words resulted in a significantly reduced response accuracy, and significantly increased response time, compared to pseudo words containing low-frequency roots. In other words, we obtained a reversed root frequency effect.

These results also allowed us to tease apart the influences of root frequency when roots are embedded in real words and in novel letter strings. The presence of high-frequency roots in real words facilitated correct responses to these words (small but significant facilitation). By contrast, in pseudo words root frequency effects appear to be reversed, as the presence of high-frequency roots made it harder for participants to correctly reject these novel strings as non-lexical items thus decreasing response accuracy and increasing accurate response time. Similarly, response times were generally significantly slower in pseudo words compared to real words, and this was especially the case for pseudo words that contained high-frequency roots.

At first glance, these results, particularly the absence of root frequency effects on lexical decision response time, may be thought of as evidence against a compulsory morphological decomposition and root identification route when processing Arabic words. As will be detailed in the General Discussion, this is not the only possible explanation for the findings, and might be an inaccurate conclusion.

General Discussion

In the current investigation, Experiment 1 aimed to replicate word frequency effects on eye movements during Arabic reading. The results obtained clearly indicated that low-frequency words attracted significantly longer fixation durations than high frequency words. Furthermore, using the same stimuli from Experiment 1, word frequency effects were obtained in a lexical decision in Experiment 3a. We explained the observed word frequency effects in Arabic in the light of the Bayesian Reader model (Norris, 2006). This model postulates that word frequency effects are obtained given the optimization of functioning of the word identification system in the linguistic environment in which it operates.

In Experiment 2, the influence of root frequency on eye movement measures was investigated during sentence reading for the first time in Arabic. Findings from previous investigations in Arabic and Hebrew, and the dual route model proposed by Frost et al. (1997) with its morphological decomposition and root identification route operating at the early stages of word identification led us to expect that root frequency would influence fixation durations. The results obtained however indicated that words containing high-frequency roots did not attract significantly shorter fixation durations during reading compared to targets containing low-frequency roots. An initial interpretation of these results would be that compulsory morphological decomposition does not happen during reading in Arabic. However, this pattern of results could also be considered in the light of the findings and theoretical account reported by Taft (2004). Taft demonstrated that under specific circumstances where readers are more likely to rely on morphological decomposition during word processing, elimination or even reversal of the influence of morphological constituents' frequency can occur. Specifically, when a high-frequency English word base (e.g., *seem*,

which is functionally similar to an Arabic root) is embedded in words of overall low orthographic frequency (e.g., *seeming*) in order to match the low orthographic frequency of a word that contains a low frequency stem like *mend* (e.g., in the word *mending*), this results in elimination or even reversal of the advantage of the high frequency base *seem* relative to *mend*. This happens because the base *seem* is considerably less frequently encountered in the continuous form *–ing*, compared to the base *mend*. Using low frequency words (e.g., *mend*) forces readers to rely on morphological decomposition of the word into base and form (see also Schreuder & Baayen, 1995), and is followed by morphological re-combination in order to complete the task at hand (e.g., sufficient identification for lexical decision). Re-combining high frequency bases with forms that are less frequent (e.g., *seem* + *–ing*) results in high processing costs that counteract the benefit of the high frequency of base *seem*. This was indeed the pattern of results reported by Taft: No significant facilitation for high frequency English word bases was found under these conditions. In the current Experiment 2, recall that due to unavoidable linguistic restrictions in Arabic, the selected stimuli featured high-frequency Arabic roots embedded in very low-frequency words, to match the low-frequency root (and word) condition. The findings reported in Experiment 2 are, thus, in line with the results reported by Taft (2004): No significant facilitation for high frequency Arabic roots was found. Importantly, adopting the account advocated by Taft, the absence of root frequency effects cannot be used to infer that morphological decomposition and root identification do not occur during early lexical processing, rather, the opposite is correct. The findings from Experiment 2 could also suggest that compulsory decomposition is followed by morphological re-combination, and the costs of re-combination, as observed in Taft's lexical decision task, generalize to silent reading.

How does the absence of significant Arabic root frequency effects compare with findings in European languages where the frequency of word morphological constituents

(lexemes) was found to influence fixation durations independently from overall compound word frequency (e.g., Hyönä & Pollatsek, 1998; Juhasz, Starr, Inhoff, & Placke, 2003)? It is likely that methodological differences can account for these different results. Specifically, when manipulating initial lexeme frequency in Finnish compounds, while attempting to keep overall word frequency constant, Hyönä and Pollatsek relied on database counts for lexeme frequency control, but on subjective ratings for overall word frequency matching. Importantly, participants' ratings indicated that both high and low frequency lexemes were embedded in Finnish words that were rated as frequently encountered and common (on a 7-point scale, 1 = highly uncommon, average ratings were 4.49 and 4.47 for the words containing high- and low-frequency lexemes respectively). A very similar procedure was employed in the investigation of lexeme frequency effects in English, with Juhasz et al. reporting target compound word commonness ratings of > 5.5 on a similar 7-point scale in all conditions where the target words contained high- or low-frequency lexemes. By contrast, the words in the current study, arguably, were all of very low orthographic frequency. It is thus possible that the processing costs at the re-combination stage for the Arabic stimuli may have been greater for high frequency roots embedded in low frequency Arabic words compared to the re-combination costs for the high frequency Finnish or English lexemes embedded in frequently encountered and common words. If so, then this may have resulted in the elimination of Arabic root frequency effects, while the Finnish and English lexeme frequency effects were preserved.

The reversed root frequency effect obtained for non-words in Experiment 3b is also in line with the findings reported by Taft (2004), and lends more support for the morphological decomposition and re-combination account discussed above. Recall that in this lexical decision experiment we used the same low-frequency real words containing high- and low-frequency roots from Experiment 2. We also decided to manipulate the root frequency

embedded in the pseudo words to allow further investigation of the role of roots in the processing of the presented letter strings. According to Taft (2004), using non-words that contain real bases in lexical decision (e.g., the base *mirth* in *mirths*, similar to the stimuli used in Experiment 3b with pseudo Arabic words containing real Arabic roots) should result in the elimination or even reversal of word base frequency effects. This is because when both real words and non-words contain real bases, the only way to discriminate between them depends on completing the morphological re-combination stage, where only real words recombine successfully with the word pattern. This, arguably, may lead to magnification of processing effects at the morphological re-combination stage. In line with this, Experiment 3b results replicated the elimination of Arabic root frequency effects in response latencies for real words. Furthermore, when pseudo words were processed, re-combination of high-frequency roots with forms that were, by definition, very unusual combinations for native readers resulted in greater processing costs. Indeed, these re-combination costs were so great that response latency and accuracy for the pseudo words containing these high-frequency roots showed a reversed root frequency effect.

Careful reading of the conclusions being drawn above invites the question: How can both the (hypothetical) presence, and absence of root frequency effects on processing time be used to argue for morphological decomposition taking place? Is this a tenable theoretical stance? As explained above, had Experiments 2 and 3b produced root frequency effects in the expected direction, an obvious conclusion would have been that such findings are in line with the operation of a morphological decomposition and root identification route, as proposed by the Frost et al. (1997)'s dual route model. Yet, the absence of these root frequency effects is presented here as supporting evidence for the operation of a morphological decomposition and root identification route. We suggest that this is, indeed, a tenable theoretical stance. To begin with, morphological decomposition is central to both the

Frost et al. model and to the account presented by Taft, this is not controversial in itself. Taft's account simply spells out what happens following morphological decomposition, and the consequences to processing of morphological decomposition and re-combination happening in a specific set of circumstances. These circumstances include those that occurred during reading of the Arabic stimuli that we selected and used in our experiments (2 and 3b).

One final related theoretical consideration remains to be discussed. On one hand, the account for processing Semitic words put forward by Frost et al. (1997) postulates a dual route model of processing. On the other hand, the theoretical account presented by Taft (2004) states that the obtained results (elimination or reversal of morphological constituent frequency effects) would be hard to accommodate in dual route accounts of morphological processing, whereas it naturally follows from obligatory morphological decomposition accounts. Thus, the question is: Are there any serious contradictions in endorsing Frost et al.'s dual route model while also claiming to support the account put forward by Taft? The likely answer is no. To begin with, Taft (p. 754) suggested that the elimination of morphological constituent frequency effects may also be accommodated by dual route accounts when pseudo word distractors contain real morphological constituents (e.g., roots or word bases) in lexical decision. Indeed, Taft concludes that "Perhaps there are differences between languages regarding the importance of the combination stage, with that stage being more important for languages that have a more productive morphology." (p. 762). This possibly applies in particular to Semitic languages where morphology is highly productive, and can potentially explain why the absence of root frequency effects was not only observed in lexical decision, but also in silent reading in Arabic (see also Schreuder & Baayen, 1995). The present findings do not allow us to speculate beyond this point, and further establishing the (in)compatibility of these two accounts (dual route vs. compulsory morphological

decomposition only) requires further comparative investigation of morphological processing in various morphological systems.

Aside from the morphological re-combination costs, it is not possible to rule out another explanation for the reversed root frequency effects obtained for pseudo words in lexical decision. These effects may have been obtained because pseudo words that contained high frequency roots were more word-like, compared to those containing the low frequency roots. Being more word-like can account for the difficulty and slowness in rejecting such novel strings, hence the reversed root frequency effect. Future investigations may be necessary to adjudicate between the morphological decomposition and re-combination costs, and the word-likeness accounts.

To summarize, our findings from eye movement measures during sentence reading, and from lexical decision response time and accuracy replicate in Arabic the widely-reported word frequency effects in silent reading and in lexical decision. Whereas the simplest explanation for the lack of root frequency effects we observed might be an absence of such effect during text reading, according to the account put forward by Taft (2004), the elimination of root frequency effects is not a sufficient argument against the operation of compulsory morphological decomposition in reading Arabic words. Rather, the results obtained are in line with previous findings that reported elimination of morphological constituent frequency effects under conditions where the processing costs of morphological re-combination can outweigh the benefits of a root being of high frequency. The reversal of the root frequency effect in pseudo words in lexical decision highlights the degree to which the costs of morphological re-combination can influence letter string processing. Our findings are the first to document word frequency and examine root frequency effects during Arabic silent sentence reading and lexical decision.

References

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, *16*, 285-311. doi:10.1080/09541440340000123

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi:10.1016/j.jml .2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. doi:10.1016/j.jml .2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67, 1-48. http://dx.doi.org/10.18637/jss.v067.i01

Bentin, S., & Frost, R. (1995). Morphological Factors in Visual Word Identification in Hebrew. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 271-292). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bertram, R. & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, *48*, 615-634. doi: 10.1016/S0749-596X(02)00539-9

Boudelaa, S. (2014). Is the Arabic mental lexicon morpheme-based or stem-based? Implications for spoken and written word recognition. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy – insights and perspectives* (pp. 31-54). New York: Springer.

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, *81*, 65-92. doi:10.1016/S0010-0277(01)00119-6

Boudelaa, S., & Marslen-Wilson, W. D. (2004). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, *90*, 106-116. doi:10.1016/S0093-934X(03)00424-3

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, *20*, 207-260. doi:10.1080/01690960444000106

Boudelaa, S., & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behavior Research Methods*, *42*, 481-487. doi:10.3758/BRM.42.2.481

Boudelaa, S., & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes*, *26*, 624-652. doi:10.1080/01690965.2010.521022

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

Boudelaa, S., Pulvermüller, F., Hauk, F., Shtyrov, Y., & Marslen-Wilson, W. D. (2009). Arabic Morphology in the Neural Language System. *Journal of Cognitive Neuroscience*, 22, 998-1010. doi:10.1162/jocn.2009.21273

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177- 178. doi:10.1038/226177a0

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, *13*, 767-782. doi:10.1016/0042-6989(73)90041-2

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

Deutsch, A., Frost, R., & Forster, K. I. (1998). Verbs and nouns are organized and accessed differently in the mental lexicon: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1238-1255. doi:10.1037/0278-7393.24.5.1238

Deutsch A, Frost, R., Pollatsek A., & Rayner, K. (2000). Early morphological effects in word recognition in Hebrew: Evidence from parafoveal preview benefit. *Language and Cognitive Processes*, *15*, 487-506. doi:10.1080/01690960050119670

Deutsch, A., Frost, R., Peleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin & Review*, *10*, 415-422. doi:10.3758/BF03196500

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. doi:10.3389/fpsyg.2014.00781

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 839-855). Oxford, England: OUP.

Drieghe, D., Brysbaert, M., & Desmet, T. (2005). Parafoveal-on-foveal effects on eye movements in text reading: Does an extra space make a difference? *Vision Research*, *45*, 1693-1706. doi:10.1016/j.visres.2005.01.010

Engbert, R., & Kliegl, R. (2011). Parallel graded attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 787-800). Oxford, England: Oxford University Press.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.

Frost, R. (2009). Reading in Hebrew vs. reading in English: Is there a qualitative difference? In K. Pugh & P. McCradle (Eds.), *How children learn to read: Current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice* (pp. 235–54). New York: Psychology Press.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*, 263-279. doi:10.1017/S0140525X11001841

Frost, R., Deutsch, A., & Forster, K. I. (2000). Decomposing complex words in a nonlinear morphology. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 751-765. doi:10.1037/0278-7393.26.3.751

Frost, R., Deutsch, A, Gilboa, O., Tannenbaum M., & Marslen-Wilson, W. D. (2000). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory & Cognition*, *28*, 1277-1288. doi:10.3758/BF03211828

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew: A masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *23*, 829-856. doi: 10.1037/0278-7393.23.4.829

Frost, R., Kugler, T., Deutsch, A., & Forster, K. I. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1293-1326. doi:10.1037/0278-7393.31.6.1293

Gwilliams, L. & Marantz, A. (2015). Non-linear processing of a linear speech stream: The influence of morphological structure on the recognition of spoken Arabic words. *Brain & Language*, *147*, 1-13. doi: 10.1016/j.bandl.2015.04.006

Haywood, J. A., & Nahmad, H. M. (1965). *A new Arabic grammar of the written language*. Surrey: Lund Humphries.

Hermena, E. W., Liversedge, S. P., & Drieghe, D. (2016). The influence of a word's number of Letters, spatial Extent, and initial bigram characteristics on eye movement control during reading: Evidence from Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 451-471. doi:10.1037/xlm0000319

Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 819-838). Oxford, England: OUP.

Hyönä, J., Bertram, R., & Pollatsek, A. (2004). Are long compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory & Cognition, 32*, 523-532. doi:10.3758/BF03195844

Hyönä, J., & Pollatsek, A. (1998). The role of component morphemes on eye fixations when reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1612-1627. doi:10.1037/0096-1523.24.6.1612

Inhoff, A.W. & Rayner, K. (1986). Parafoveal word processing during eye fixation in reading: Effects of word frequency. *Perception & Psychophysics, 40*, 431-439. doi:10.3758/BF03208203

Juhasz, B. G. (2008). The processing of compound words in English: Effects of word length on eye movements during reading. *Language and Cognitive Processes*, *23*, 1057-1088. doi:10.1080/01690960802144434

Juhasz, B. G., Starr, M., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decision and eye fixations. *British Journal of Psychology*, *94*, 223-244. doi:10.1348/000712603321661903

Juhasz, B. J., & Pollatsek, A. (2011). Lexical influences on eye movements in reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 873-893). Oxford, England: OUP.

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, *62*, 83-97. doi:10.1016/j.jml.2009.10.001

Luke, S. G., & Christianson, K. (2011). Stem and whole-word frequency effects in the processing of inflected verbs in and out of sentence context. *Language and Cognitive Processes*, *26*, 1173-1192. doi: 10.1080/01690965.2010.510359

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9. 12-2.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165-178. doi:10.1037/h0027366

Niswander-Klement, E., & Pollatsek, A. (2006). The effect of root frequency, word frequency, and length on the processing of prefixed words during reading. *Memory and Cognition*, *34*, 685-702. doi:10.3758/BF03193588

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*,*113*, 327-357. doi:10.1037/0033-295X.113.2.327

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, *137*, 434-455. doi:10.1037/a0012799

Perea, M., Abu Mallouh, R., & Carreiras, M. (2010). The search for an input-coding scheme: Transposed-letter priming in Arabic. *Psychonomic Bulletin & Review*, *17*, 375-380. doi:10.3758/PBR.17.3.375

Prunet, J. F., Béland, R., & Idrissi, A. (2000). The mental representation of Semitic words. *Linguistic Inquiry*, *31*, 609-648. doi:10.1162/002438900554497

R Core Team. (2016). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http:// www.R-project.org/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457-1506. doi:10.1080/17470210902816461

Reichle, E. D. (2011). Serial attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 767-786). Oxford, England: Oxford University Press.

Schultz, E. (2004). *A student grammar of Modern Standard Arabic*. Cambridge: CUP.

Schreuder, R., & Baayan, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131-154). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, *57A*, 745-765. doi:10.1080/02724980343000477

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020-2045. doi:10.1037/xge0000014

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison

using 855 *t* tests. *Perspectives on Psychological Science*, *6*, 291-8. doi:10.1177/1745691611406923.

Table 1

Descriptive Statistics of Eye Movement Measures at Pre-Target and Target Regions Calculated Across Subjects (Experiment 1)

Region	Eye Movement Measure	High Orthographic Frequency Mean (SD)	Low Orthographic Frequency Mean (SD)
Pre-Target	Last Fixation Duration in First Pass	277 (110)	269 (110)
	Gaze Duration (ms)	332 (163)	345 (215)
Target	Skipping	0.02 (0.1)	0.03 (0.2)
	First Pass Fixation Count	1.5 (0.7)	1.8 (1.0)
	First Fixation (ms)	270 (104)	305 (133)
	Single Fixation (ms)	289 (107)	335 (142)
	Gaze Duration (ms)	375 (178)	510 (302)
	Go Past (ms)	449 (312)	624 (387)

Table 2

Descriptive Statistics of Eye Movement Measures at Pre-Target and Target Regions Calculated Across Subjects (Experiment 2)

Sentence Region	Eye Movement Measure	High Root Frequency Mean (SD)	Low Root Frequency Mean (SD)
Pre-Target	Last Fixation Duration in First Pass	282 (110)	270 (105)
	Gaze Duration (ms)	340 (169)	349 (177)
Target	Skipping	0.10 (0.3)	0.13 (0.3)
	First Pass Fixation Count	1.5 (0.7)	1.5 (0.8)
	First Fixation (ms)	300 (130)	303 (123)
	Single Fixation (ms)	321 (131)	323 (124)
	Gaze Duration (ms)	440 (238)	431 (238)
	Go Past (ms)	525 (316)	533 (365)

Table 3

Descriptive Statistics of Lexical Decision Measures Calculated Across Subjects (Experiment 3a)

Measure	Word - High Orthographic Frequency Mean (SD)	Word - Low Orthographic Frequency Mean (SD)	Pseudo Word Mean (SD)
Accuracy (%)	99.5 (0.1)	90.5 (0.3)	93.1 (0.3)
Reaction Time (ms)	700 (183)	870 (248)	929 (269)

Table 4

Descriptive Statistics of Lexical Decision Measures Calculated Across Subjects (Experiment 3b)

Measure	Word - High Root Frequency Mean (SD)	Word - Low Root Frequency Mean (SD)	Pseudo Word - High Root Frequency Mean (SD)	Pseudo Word - Low Root Frequency Mean (SD)
Accuracy (%)	94.4 (0.2)	90.2 (0.3)	85.8 (0.3)	90.6 (0.3)
Reaction Time (ms)	829 (233)	849 (233)	973 (250)	909 (264)

Figure 1

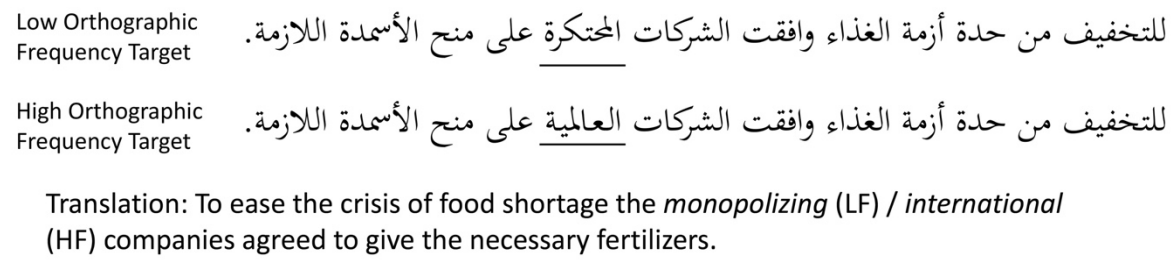

Figure 1. Sample stimuli for Experiment 1. The target words are underlined in Arabic, and italicized in the translation.

Figure 2

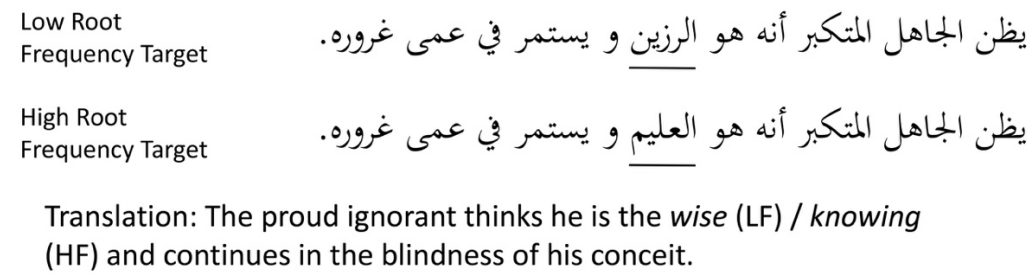

Figure 2. Sample stimuli for Experiment 2. The target words are underlined in Arabic, and italicized in the translation.

Figure 3

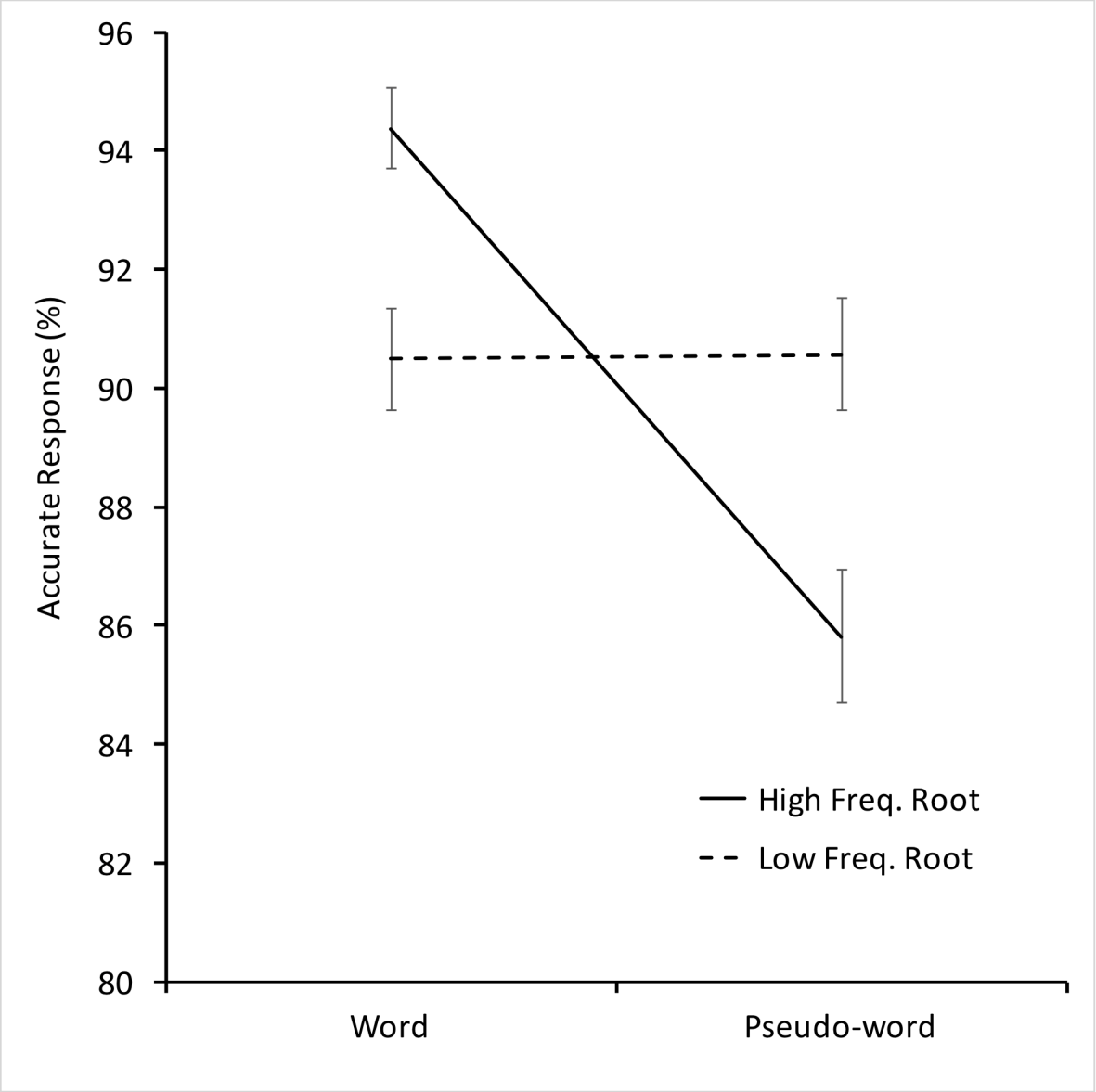

Figure 3. Lexical decision accuracy (Experiment 3b). Interaction between target lexicality and root frequency. The error bars represent the standard error.

Figure 4

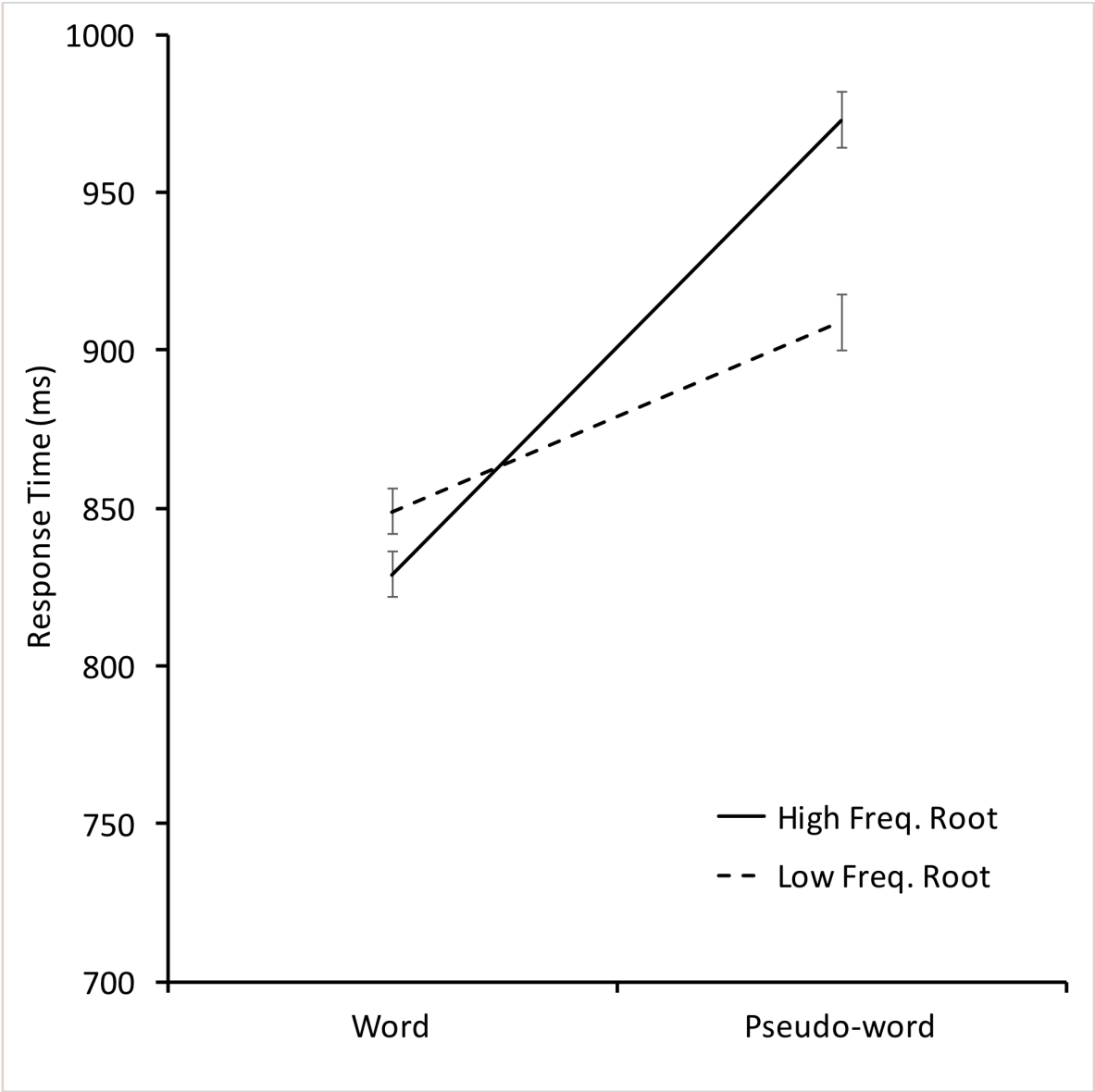

Figure 4. Lexical decision reaction time (Experiment 3b). Interaction between target lexicality and root frequency. The error bars represent the standard error.

Appendix 1

Stimuli for Experiment 1

Word Orthographic Frequency Manipulation

Sentence	Target Syntactic Case	Word Frequency / Item Identifier
في لقاء حافل للفنانين التشكيليين **الخزافين** قدم الفنان أحمد شكري عمله الجديد.	n. pl. m.	LowFreq_01
في لقاء حافل للفنانين التشكيليين **المصريين** قدم الفنان أحمد شكري عمله الجديد.	n. pl. m.	HighFreq_01
وضحت الوثيقة تاريخ **السحوبات** المالية من حساب الشركة في الخمس أعوام الأخيرة.	n. pl. f.	LowFreq_02
وضحت الوثيقة تاريخ **المؤسسات** المالية التي تسببت في الانهيار الاقتصادي.	n. pl. f.	HighFreq_02
حاول الفريق لساعات طويلة إصلاح **الغلايات** في المولد الكهربائي بلا نجاح.	n. pl. f.	LowFreq_03
حاول الفريق لساعات طويلة إصلاح **العلاقات** في مؤتمر المنظمة بلا نجاح.	n. pl. f.	HighFreq_03
فصلت الصحف ملابسات القبض على **الجاسوسة** التي حاولت نشر الفتن في العاصمة.	n. s. f.	LowFreq_04
فصلت الصحف ملابسات القبض على **المجموعة** التي حاولت نشر الفتن في العاصمة.	n. s. f.	HighFreq_04
وضح المراسل تفاصيل غرق المركبة **المنكودة** في البحر المتوسط مساء أمس.	adj. s. f.	LowFreq_05
وضح المراسل تفاصيل غرق المركبة **العسكرية** في البحر المتوسط مساء أمس.	adj. s. f.	HighFreq_05
قبضت الشرطة على مستوردي المنتجات الغذائية **المتعفنة** التي تسببت في تسمم الأطفال.	adj. s. f.	LowFreq_06
قبضت الشرطة على مستوردي المنتجات الغذائية **الأجنبية** منتهية الصلاحية.	adj. s. f.	HighFreq_06
دقق القادة النظر في خطط **السالفين** لمكافحة الجهل والفقر والمرض.	n. pl. m.	LowFreq_07
دقق القادة النظر في خطط **المستقبل** لمكافحة الجهل والفقر والمرض.	n. pl. m.	HighFreq_07
توجهت كافة مجهودات الشعب **الكفاحية** لمحاربة الظلم وانعدام المساواة المادية.	adj. s. f.	LowFreq_08
توجهت كافة مجهودات الشعب **الرئيسية** لمحاربة الظلم وانعدام المساواة المادية.	adj. s. f.	HighFreq_08
ندد المحللون والمراقبون بنتائج الانقسام **القبائلي** الذي بدد ثروات البلاد.	adj. s. m.	LowFreq_09
ندد المحللون والمراقبون بنتائج الانقسام **المتزايد** الذي حال دون إنهاء الصعوبات الاقتصادية.	adj. s. m.	HighFreq_09
خشي المسؤولون من تدهور حالة البلاد **المتوقدة** بالغضب على الأحوال الاقتصادية.	adj. s. f.	LowFreq_10
خشي المسؤولون من تدهور حالة البلاد **الداخلية** المتوترة وخاصة حالة الاقتصاد.	adj. s. f.	HighFreq_10
حاول السفراء إصلاح العلاقات **المتموجة** مع بقية دول المنطقة.	adj. s. f.	LowFreq_11
حاول السفراء إصلاح العلاقات **الخارجية** مع بقية دول المنطقة.	adj. s. f.	HighFreq_11
قال المنتقدون أن محاولات الوساطة **المشلولة** كانت غير مثمرة بسبب تعنت الطرف الأخر.	adj. s. f.	LowFreq_12
قال المنتقدون أن محاولات الوساطة **السياسية** كانت غير مثمرة بسبب تعنت الطرف الأخر.	adj. s. f.	HighFreq_12
بعد كارثة تسرب النفط قدمت الشركات **المنتفعة** كل المساعدات المطلوبة لإنقاذ البيئة.	adj. s. f.	LowFreq_13
بعد كارثة تسرب النفط قدمت الشركات **المتخصصة** كل المساعدات المطلوبة لإنقاذ البيئة.	adj. s. f.	HighFreq_13
للتخفيف من حدة أزمة الغذاء وافقت الشركات **المحتكرة** على تسويق الأسمدة اللازمة.	adj. s. f.	LowFreq_14
للتخفيف من حدة أزمة الغذاء وافقت الشركات **العالمية** على تسويق الأسمدة اللازمة.	adj. s. f.	HighFreq_14
وجدت قوات الإنقاذ الفتاة **المقتولة** والعديد من الضحايا الأخرين بعد العمل الإرهابي.	adj. s. f.	LowFreq_15
وجدت قوات الإنقاذ الفتاة **المعروفة** والعديد من الضحايا الأخرين بعد العمل الإرهابي.	adj. s. f.	HighFreq_15
إحتج عدد كبير من العملاء بسبب الأخطاء **المطبعية** في النصوص و الرسوم المطبوعة.	adj. s. f.	LowFreq_16
إحتج عدد كبير من العملاء بسبب الأخطاء **الصناعية** التي أثرت سلبا على جودة المنتجات.	adj. s. f.	HighFreq_16

LowFreq_17	n. pl. m.	في خلال جولته شجع سيادته الرئيس جميع **المحيطين** بموكبه على الأمل والعمل.
HighFreq_17	n. pl. m.	في خلال جولته شجع سيادته الرئيس جميع **اللاعبين** على الأمل والعمل.
LowFreq_18	n. pl. f.	صرحت مصادر قيادية أن **المقذوفات** قد بلغت أهدافها بنجاح ودقة.
HighFreq_18	n. pl. f.	صرحت مصادر قيادية أن **المحادثات** قد بلغت أهدافها بنجاح.
LowFreq_19	n. pl. f.	وضح التقرير قدر أرباح الشركة بعد نجاح **المخبوزات** الطازجة ومنتجاتها الغذائية الأخرى.
HighFreq_19	n. pl. f.	وضح التقرير قدر أرباح الشركة بعد نجاح **المشروعات** السكنية التي أتمتها العام الماضي.
LowFreq_20	adj. s. f.	بسبب عدم موافقة الصين على السياسات **التصديرية** اليابانية لم يصل الأطراف إلى أي تفاهم.
HighFreq_20	adj. s. f.	بسبب عدم موافقة الصين على السياسات **التنفيذية** الجديدة لم يصل الأطراف إلى أي تفاهم.
LowFreq_21	adj. s. f.	كان الضيف الألماني شغفا بالعمارة **الإبداعية** في منطقة الجيزة.
HighFreq_21	adj. s. f.	كان الضيف الألماني شغفا بالعمارة **الإسلامية** في القاهرة والإسكندرية.
LowFreq_22	n. pl. f.	بدا الإرتياح على أوجه الجميع إذ انتهت **الإشكالات** بحصول كل الأطراف على تسوية مقبولة.
HighFreq_22	n. pl. f.	بدا الإرتياح على أوجه الجميع إذ انتهت **المفاوضات** بحصول كل الأطراف على تسوية مقبولة.
LowFreq_23	n. pl. m.	في صباح الإثنين الماضي كان **الممتحنين** في حالة قلق بسبب تأخر أوراق الأسئلة.
HighFreq_23	n. pl. f.	في صباح الإثنين الماضي كان **المواطنين** في حالة قلق بسبب تأخر نتائج الانتخابات.
LowFreq_24	adj. s. f.	كانت الأسرة سعيدة بلقاء المعلمة **المتحمسة** في مساء الجمعة الماضية.
HighFreq_24	adj. s. f.	كانت الأسرة سعيدة بلقاء المعلمة **المطلوبة** في مساء الجمعة الماضية.
LowFreq_25	n. pl. f.	علقت الحكومة على **التعقيبات** التي أصدرها وزير الخارجية البريطاني ووصفتها بعدم الدقة.
HighFreq_25	n. pl. f.	علقت الحكومة على **المعلومات** التي أصدرها وزير الخارجية البريطاني ووصفتها بعدم الدقة.
LowFreq_26	adj. s. f.	استفاد كل الطلاب من النصوص **الإيضاحية** الإضافية التي ذكرها المعلم في نهاية الدرس.
HighFreq_26	adj. s. f.	استفاد كل الطلاب من النصوص **التاريخية** الإضافية التي ذكرها المعلم في نهاية الدرس.
LowFreq_27	n. pl. m.	قبضت الشرطة صباح الأمس على جميع **المختبئين** بالمباني المهجورة في المنطقة.
HighFreq_27	n. pl. m.	قبضت الشرطة صباح الأمس على جميع **المسؤولين** المتهمين بالرشوة والفساد.
LowFreq_28	adj. f. nA.	صفق الجميع لفوز السباحة الأولمبية **البرازيلية** في المسابقة الأخيرة.
HighFreq_28	adj. f. nA.	صفق الجميع لفوز السباحة الأولمبية **الفلسطينية** في المسابقة الأخيرة.
LowFreq_29	adj. m. nA.	من الواضح أن المنتج **السنغافوري** كان رديء الجودة مقارنة بمنتجنا الوطني.
HighFreq_29	adj. m. nA.	من الواضح أن المنتج **الإسرائيلي** كان رديء الجودة مقارنة بمنتجنا الوطني.
LowFreq_30	n. m. nA.	وضح الكاتب كيف أن الفكر السياسي **الأوتوقراطي** لا يخلوا من التناقضات والمشاكل.
HighFreq_30	n. m. nA.	وضح الكاتب كيف أن الفكر السياسي **الديموقراطي** لا يخلوا من التناقضات والمشاكل.

Note. Target word pairs are in boldface. n. = noun; adj. = adjective; f. = feminine; m. = masculine, s. = singular; pl. = plural; nA. = Arabized word of non-Arabic origin.

Appendix 2

Stimuli for Experiment 2

Word Orthographic Frequency Manipulation

Sentence	Target Syntactic Case	Root Frequency / Item Identifier
كان من نصيب الرجل المرض النفسي **والجذام** ومع ذلك لم يشتكي.	n. s. m.	LowFreq_01
كان من نصيب الرجل المرض النفسي **والأنين** ومع ذلك لم يشتكي.	n. s. m.	HighFreq_01
في صباح الإثنين قابل التاجر **الباعة** لمناقشة سياسة البيع وأسعار الشراء.	n. pl. m.	LowFreq_02
في صباح الإثنين قابل التاجر **العميل** لمناقشة سياسة البيع وأسعار الشراء.	n. s. m.	HighFreq_02
يظن الجاهل المتكبر أنه هو **الرزين** ويستمر في عمى غروره.	adj. s. m.	LowFreq_03
يظن الجاهل المتكبر أنه هو **العليم** ويستمر في عمى غروره.	adj. s. m.	HighFreq_03
قامت القوات بالتخلص من جميع الوحدات **المعطوبة** بسلاح الطيران.	adj. s. f.	LowFreq_04
قامت القوات بالتخلص من جميع الوحدات **المتآمرة** بسلاح المشاة و سلاح المدفعية.	adj. s. f.	HighFreq_04
وصف الجميع كيف كان مظهره **المنمق** متوافقا مع طبيعته و شخصيته.	adj. s. m.	LowFreq_05
وصف الجميع كيف كان مظهره **القويم** متوافقا مع طبيعته و شخصيته.	adj. s. m.	HighFreq_05
من أخطر العادات **التسكع** في الشوارع بغير علم و لا هدف.	n. s. m.	LowFreq_06
من أخطر العادات **التقول** في حقوق الأخرين بغير علم و لا هدف.	n. s. m.	HighFreq_06
وصفت القصة كيف أن تجمع **الرعاع** لم يؤدي إلى حل المشاكل.	n. pl. m.	LowFreq_07
وصفت القصة كيف أن تجمع **العوام** لم يؤدي إلى حل المشاكل.	n. pl. m.	HighFreq_07
صورت الكاميرا كيف بدا الماء وكأنه **يتباطأ** في النهر بعد ذوبان الثلج.	v. pr. s. m.	LowFreq_08
صورت الكاميرا كيف بدا الماء وكأنه **يتدافع** في سباق بعد ذوبان الثلج.	v. pr. s. m.	HighFreq_08
في قاعات المحاكم **يتعامى** بعض المحامين عن بعض مبادئ القانون.	v. pr. s. m.	LowFreq_09
في قاعات المحاكم **يترافع** بعض المحامين مهملين بعض مبادئ القانون تعمدا.	v. pr. s. m.	HighFreq_09
لم يكن التلميذ **مستوعب** فساعده المعلم على تفهم القواعد اللغوية.	adj. s. m.	LowFreq_10
لم يكن التلميذ **متجاوب** فساعده المعلم على تفهم أهمية المشاركة في الدرس.	adj. s. m.	HighFreq_10
وصفت الصحافة إجابات الرئيس بأنها **مبعثرة** وغير موحدة القصد.	adj. s. f.	LowFreq_11

HighFreq_11	adj. s. f.	وصفت الصحافة إجابات الرئيس بأنها **مبرمجة** وغير مشجعة على الحوار.
LowFreq_12	adj. s. f.	وصف البعض قوات الأمن بأنها **مطمئنة** مما قد يؤدي إلى تهدئة الأوضاع.
HighFreq_12	adj. s. f.	وصف البعض قوات الأمن بأنها **مسيطرة** على الموقف مما قد يؤدي إلى تهدئة الأوضاع.
LowFreq_13	adj. s. f.	تم الكشف عن رموز **رمادية** باهتة داخل الكهف لا يمكن قراءتها بعد.
HighFreq_13	adj. s. f.	تم الكشف عن رموز **كتابية** باهتة داخل الكهف لا يمكن قراءتها بعد.
LowFreq_14	adj. s. f.	تداول الكاتب بشكل مشوق حقائق **بديهية** في كتابه الجديد عن الاقتصاد العالمي.
HighFreq_14	adj. s. f.	تداول الكاتب بشكل مشوق حقائق **تقدمية** في كتابه الجديد عن الاقتصاد العالمي.
LowFreq_15	adj. s. f.	ألقى المرشح خطابات **منبرية** واسعة التأثير على الشعب قبل الإنتخابات.
HighFreq_15	adj. s. f.	ألقى المرشح خطابات **تآمريه** واسعة التأثير على الشعب قبل الإنتخابات.
LowFreq_16	adj. s. m.	إنصرف الموظف عن الإجتماع **متشائما** و مهرول الخطى.
HighFreq_16	adj. s. m.	إنصرف الموظف عن الإجتماع **متعاليا** و مهرول الخطى.
LowFreq_17	adj. s. f.	إنتهت القصة بمشهد رقدت فيه المرأة **السقيمة** على فراش موتها طالبة غفران السماء.
HighFreq_17	adj. s. f.	إنتهت القصة بمشهد رقدت فيه المرأة **الفاترة** على فراش موتها طالبة غفران السماء.
LowFreq_18	adj. s. m.	الرياح الجارفة والمطر **المنهمر** يهددان المحاصيل بالفناء.
HighFreq_18	adj. s. m.	الرياح الجارفة والمطر **المقترب** يهددان المحاصيل بالفناء.
LowFreq_19	adj. s. m.	يسخر البعض من الرجل **المحتشم** لأنه ليس مثلهم.
HighFreq_19	adj. s. m.	يسخر البعض من الرجل **المعطاء** لأنه ليس مثلهم.
LowFreq_20	adj. pl. m.	تبحث الشركة عن موظفين **محتاجين** للتعيين بوظائف مؤقتة.
HighFreq_20	adj. pl. m.	تبحث الشركة عن موظفين **متعلمين** للتعيين بوظائف مؤقتة.
LowFreq_21	adj. s. f.	هددت المشكلات **الجاثمة** على صدر الوطن بتفكيك وحدة الشعب.
HighFreq_21	adj. s. f.	هددت المشكلات **المتقدة** على صدر الوطن بتفكيك وحدة الشعب.
LowFreq_22	adj. s. f.	حذر عمال الإنقاذ السكان إذ تسربت الغازات **الخامدة** بعد ثورة البركان الأخيرة.
HighFreq_22	adj. s. f.	حذر عمال الإنقاذ السكان إذ تسربت الغازات **المؤذية** بعد ثورة البركان الأخيرة.
LowFreq_23	adj. s. f.	ناقش البرنامج التلفزيوني قضايا **ترفيهية** وكأنها في غاية الأهمية.
HighFreq_23	adj. s. f.	ناقش البرنامج التلفزيوني قضايا **مجتمعية** وكان النقاش مفيدا.
LowFreq_24	adj. s. f.	وصفت الجماهير تلك القرارات بأنها مهزلة **تحريضية** إستهدفت تدمير الفريق القومي.
HighFreq_24	adj. s. f.	وصفت الجماهير تلك القرارات بأنها مهزلة **تحكيمية** إستهدفت تدمير الفريق القومي.

LowFreq_25	adj. s. f.	واجه دعاة الإنقلاب مشكلات **مستعصية** حالت دونهم ودون أهدافهم.
HighFreq_25	adj. s. f.	واجه دعاة الإنقلاب مشكلات **توحيدية** إذ لم يمكنهم الإتفاق على الأهداف.
LowFreq_26	adj. s. f.	بعد الحادثة المروعة تبنى الرجل نظرة **تشاؤمية** على الحياة ولم يعد يكترث بشيء.
HighFreq_26	adj. s. f.	بعد الحادثة المروعة تبنى الرجل نظرة **تحويلية** على الحياة وقرر أن يغير حياته.
LowFreq_27	adj. s. f.	كان الرجل مهتما بدراسة الأنظمة **الفقهية** وعثر على كتب متخصصة ومفيدة.
HighFreq_27	adj. s. f.	كان الرجل مهتما بدراسة الأنظمة **العددية** في الحضارات القديمة وعثر على بعض الكتب المتخصصة.
LowFreq_28	adj. s. f.	قال المحقق في القضية أن المناقشات **الشفهية** لا تفيد و أن المطلوب هو أدلة واضحة و مكتوبة.
HighFreq_28	adj. s. f.	قال المحقق في القضية أن المناقشات **القدرية** لا تفيد و أن المطلوب هو أدلة واضحة.
LowFreq_29	adj. s. f.	صفحات التاريخ ممتلئة بتفاصيل الصراعات **الوحشية** بين مجموعات البشر ومازال البشر يتناحرون.
HighFreq_29	adj. s. f.	صفحات التاريخ ممتلئة بتفاصيل الصراعات **القبلية** بين مجموعات البشر ومازال البشر يتناحرون.
LowFreq_30	n. du. m.	في صباح اليوم حطم الجنود **مدرعتين** في المنطقة الأثرية المحمية مما أثار غضب بعض المثقفين.
HighFreq_30	n. du. m.	في صباح اليوم حطم الجنود **تمثالين** في المنطقة الأثرية المحمية مما أثار غضب بعض المثقفين.

Note. Target word pairs are in boldface. n. = noun; adj. = adjective; f. = feminine; m. = masculine, s. = singular; du. = dual; pl. = plural; v. = verb; pr. = present tense.

Appendix 3

Stimuli for Experiments 3a and 3b

Experiment 3a

Pseudo words	Low-Frequency Word	Pseudo words	High-Frequency Word	Item #
الخزاكين	الخزافين	القسغيين	المصريين	1
السخوذات	السحوبات	المؤشغات	المؤسسات	2
الهلايات	الغلايات	العلاغات	العلاقات	3
الجاشوزة	الجاسوسة	المجبوغة	المجموعة	4
الممكوذة	المنكودة	العبكلية	العسكرية	5
المتغفتة	المتعفنة	الغرنشية	الأجنبية	6
السالضين	السالفين	المستقبخ	المستقبل	7
الكتاسية	الكفاحية	الرليقية	الرئيسية	8
القبائفي	القبائلي	الأيريعي	المتزايد	9
المتوجشا	المتوقدة	الداجعية	الداخلية	10
المتموبا	المتموجة	الخادثية	الخارجية	11
الملنونة	المشلولة	القياجية	السياسية	12
المتغبكة	المنتفعة	الشعوكية	المتخصصة	13
المحتلرة	المحتكرة	الغالزية	العالمية	14
المفنوغة	المقتولة	الغلاشية	المعروفة	15
المطبفية	المطبعية	الضماعية	الصناعية	16
المحفخين	المحيطين	اللافغين	اللاعبين	17
المقسوضات	المقذوفات	المجادقات	المحادثات	18
المخبوفات	المخبوزات	المشتوصات	المشروعات	19
التصفيجية	التصديرية	الأكشيكية	التنفيذية	20
القرغوفية	الإبداعية	الإصقاشية	الاسلامية	21

22	المفاوضات	المغاوعات	الإشكالات	الإشطاقات
23	المواطنين	المواظعين	الممتحنين	الممتخثين
24	المطلوبة	اللدفافية	المتحمسة	البغلشيية
25	المعلومات	المصهوشات	التعقيبات	التغفيشات
26	التاريخية	التاضيكية	الإيضاحية	الإيغاكية
27	المسؤولين	المسأشيين	المختبئين	المختسطين
28	الفلسطينية	الفسغالية	البرازيلية	البقاريثية
29	الإسرائيلي	الإسراشيات	السنغافوري	السنغقاتية
30	الديموقراطي	السيكوشبالي	الاوتوقراطي	الأوغروباظي

Experiment 3b

Pseudo words	Low-Frequency Root	Pseudo words	High-Frequency Root	Item #
الحوقت	الجذام	العملل	الأنين	1
الزعفت	الباعة	الجمعن	العميل	2
العشمت	الرزين	كمعلمت	العليم	3
المجنوغة	المعطوبة	المتننفبة	المتآمرة	4
يتزنقت	المنمق	الدولت	القويم	5
يتسخمت	التسكع	اليرأس	التقول	6
المخطت	الرعاع	يتمثلت	العوام	7
يتفلجت	يتباطأ	يتعادد	يتدافع	8
يتججشت	يتعامى	يتحولت	يترافع	9
مسيقصل	مستوعب	متقومن	متجاوب	10
يزمجرة	مبعثرة	يبرمجت	مبرمجة	11
يزعفرت	مطمئنة	يسيطرت	مسيطرة	12
متحسكت	رمادية	قبيلين	كتابية	13
ستفغرت	بديهية	تقولوت	تقدمية	14
مبزقية	منبرية	يتبينت	تآمرية	15
متخامصا	متشائما	متسالمت	متعاليا	16
الخويصت	السقيمة	المعومت	الفاترة	17
المنخزم	المنهمر	المكللت	المقترب	18
المثملن	المحتشم	التقدار	المعطاء	19
مدامثين	محتاجين	متبللوت	متعلمين	20
الفانرة	الجاثمة	الينسبا	المتقدة	21
الساحمت	الخامدة	التمصرت	المؤذية	22
تربيقين	ترفيهية	تكتيبان	مجتمعية	23
تشهيمية	تحريضية	يتقديمت	تحكيمية	24

25	توحيدية	ميجددون	مستعصية	متخطلون
26	تحويلية	يتعربية	تشاؤمية	تجاوخية
27	العددية	الحققيت	الفقهية	الهمريت
28	القدرية	الشكلون	الشفهية	الربغية
29	القبلية	الشركون	الوحشية	التزربي
30	تمثالين	يتوحدين	مدرعتين	متبلتيت